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

Satyakam Dash1, Daniel G. Olson2, Siu Hung Joshua Chan3, Daniel Amador-Noguez4, Lee R. Lynd2 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

4Department of Bacteriology, University of Wisconsin, Madison, Wisconsin, USA

E-mail: Satyakam Dash: satyakam@psu.edu, Daniel G. Olson: Daniel.G.Olson@dartmouth.edu, Siu Hung Joshua Chan: joshua.chan@colostate.edu, Daniel Amador-Noguez : amadornoguez@wisc.edu, Lee R. Lynd : lee.r.lynd@dartmouth.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 maximum ethanol titer achieved to date remains below industrially required targets. Several studies have analyzed the impact of increasing ethanol concentration on *C. thermocellum*’s membrane properties, cofactor pool ratios, and altered enzyme regulation. In this study, we explore the extent to which thermodynamic equilibrium limits maximum ethanol titer.

**Results**

We used the max-min driving force (MDF) algorithm [1] to identify the range of allowable metabolite concentrations that maintain a negative free energy of change for all reaction steps in the pathway from cellobiose to ethanol. To this end, we used a time-series metabolite concentration dataset to flag glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as the key thermodynamic bottleneck under high external ethanol concentrations. Thermodynamic analysis was also deployed in a prospective mode to evaluate genetic interventions which can improve pathway thermodynamics by generating minimal set of reactions or elementary flux modes (EFMs) which possess unique genetic variations while ensuring mass and redox balance with ethanol production. MDF evaluation of all generated (336) EFMs identifies that, i) pyruvate phosphate dikinase (PPDK) has a higher pathway MDF than the malate shunt alternative due to limiting CO2 concentrations under physiological conditions, and ii) NADPH dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN) can alleviate thermodynamic bottlenecks at high ethanol concentrations. The combination of ATP linked phosphofructokinase (PFK-ATP) and NADPH linked alcohol dehydrogenase (ADH-NADPH) with NADPH linked aldehyde dehydrogenase (ALDH-NADPH) or ferredoxin: NADP+ oxidoreductase (NADPH-FNOR) emerges as the best intervention strategy for ethanol production with a similar MDF but higher ATP generation than the pathway employed by the ethanologen *T. saccharolyticum*.

**Conclusions**

Expanding the list of measured intracellular metabolites and improving the quantification accuracy of measurements can directly impact the fidelity of pathway thermodynamics analysis in *C. thermocellum*. This study demonstrates even before addressing an organism’s enzyme kinetics and allosteric regulations, pathway thermodynamics can flag pathway bottlenecks and propose testable strategies for restoring pathway thermodynamic feasibility.

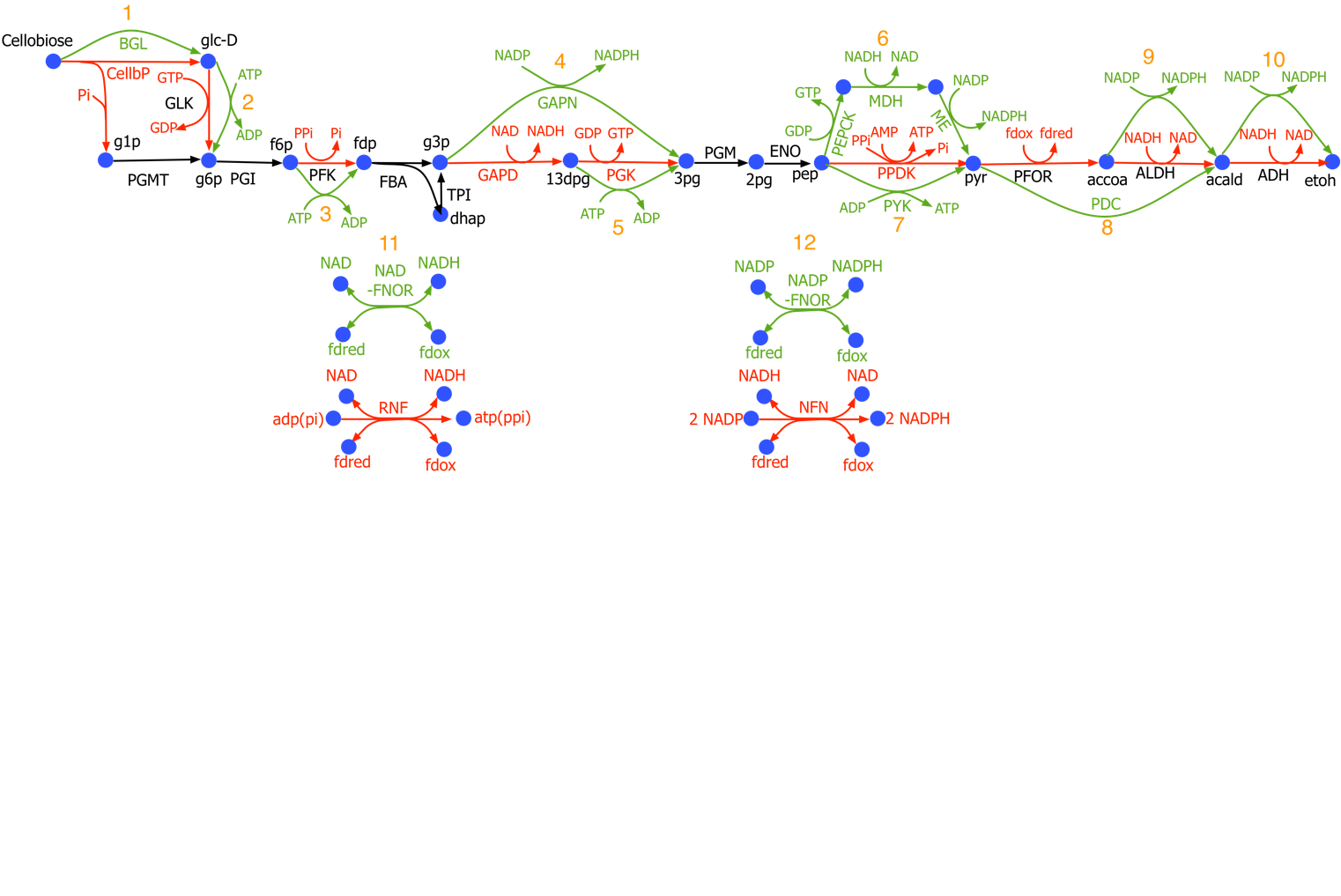
**Background**

Biofuels have emerged as a promising alternative for reducing our dependency on fossil fuels [2, 3]. 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 [4]. *Clostridium thermocellum* is a promising candidate as it can both solubilize cellulose and produce ethanol. However, low ethanol titers [5], linked to ethanol tolerance [6, 7], have so far prevented its wide-spread industrial adoption. Several studies have delved into the problem of ethanol tolerance in regards to change in membrane composition, varying cofactor pool ratios, and feedback regulations which affect enzyme kinetics [7-10]. An alternate cause for the low ethanol titer problem explored in this study involves identifying and modifying reaction steps that are bottlenecked because of thermodynamic limitations.

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 [11]. Kinetic parameterization procedures often use thermodynamic imperatives (i.e., negative free energy) to reduce the search space of feasible kinetic parameters and restrict reaction reversibility [12]. Thermodynamic analysis can be extended beyond single reactions to entire pathways. The thermodynamic feasibility of an entire multi-step linear pathway can be evaluated by identifying the reaction step with the smallest (though still negative) change in Gibbs free energy (ΔrG′). The negative of the smallest ΔrG′ (thus a positive quantity) is referred to as the minimum driving force (MDF) denoting the thermodynamic bottleneck of the pathway. The max-min driving force formulation [1] identifies the maximum MDF achievable in the pathway by modulating the metabolite concentrations within physiological limits [1]. The variability in metabolite concentrations under the maximum MDF can also be used to assess the impact of metabolite pool accumulations and/or depletions on pathway feasibility [1]. Thermodynamic bottleneck analysis has already been applied to study the cause of growth cessation in *C. thermocellum* under high substrate loading conditions, suggesting that fermentation is inhibited at the pyruvate to acetyl-coA conversion step due to accumulation of hydrogen (which inhibits the pyruvate formate oxidoreductase (PFOR) reaction via increased reduced ferredoxin levels) and formate (which inhibits the pyruvate formate lyase (PFL) reaction) [9].

# Results and discussion

We studied the impact of increasing ethanol concentration on *C. thermocellum*’s metabolism using the MDF method of analysis. We restricted our analysis to glycolysis along with ethanol production pathway (Figure 1) and its cofactor regeneration system. Using thermodynamic feasibility of time varying metabolite concentrations for *C. thermocellum* grown with and without external ethanol addition, thermodynamically inconsistent measurements are pinpointed for dihydroxyacetone phosphate (dhap) and fructose bisphosphate (fdp). The amended metabolite concentration dataset (excluding dhap and fdp) is subsequently used to constrain the wild-type pathway thermodynamics of *C. thermocellum.* It revealed that at high ethanol concentrations glyceraldehyde-3-phosphate dehydrogenase (GAPDH) becomes the thermodynamic bottleneck (i.e., MDF = -0.22 kJ/mol) due to rising NADH levels thus seizing ethanol production. In addition, twelve plausible metabolic interventions by modifying cofactor dependencies of pathway enzymes (see Table 1) are computationally explored to restore positivity of MDF at high ethanol concentrations. Notably, several of these genetic interventions had already been explored in *C. thermocellum* with reported increase in ethanol titers [13-16], however the mechanism was not always fully known. These enzyme modifications, both individually and in combination are evaluated by using the concept of elementary flux modes (EFMs). An EFM is defined as a redox and energy balanced minimal set of reactions that under steady-state conditions allows for the generation of set of products from a given set of reactants [17]. Here we exhaustively identified all EFMs that span the production of ethanol from cellobiose while superimposing onto the set of allowable reactions all additional pathway bypasses and enzyme cofactor preference modifications. We evaluated the MDF for 336 EFMs spanning all possible combinations of pathway modifications at 1 M external ethanol concentration (Supplementary Table XX). Incorporation of the malate shunt (by replacing PPDK) leads to the lowest MDF (i.e., -0.13 kJ/mol) due to constraints by intracellular CO2 and oxaloacetate (oaa) concentrations alluding to supersaturation of CO2 (increases substrate concentration) or instability of oaa (depletes product pool) to ensure malate shunt feasibility in WT *C. thermocellum*. In contrast, replacing GAPDH and PGK with the NADP-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPN) significantly improves the pathway MDF (i.e., 8.73 kJ/mol). This modification circumvents the thermodynamic bottleneck at GAPDH (by replacing NADH with NADPH) at the expense of reduced ATP generation (from 6 to 2 mol ATP/mol cellobiose). The most efficient genetic intervention involves changing the cofactor association of PFK to ATP and alcohol dehydrogenase (ADH) reaction to NADPH along with either NADPH linked aldehyde dehydrogenase(ALDH) or ferredoxin: NADP+ oxidoreductase (NADPH-FNOR) to maintain a positive MDF (i.e., 4.32 kJ/mol) similar to the high ethanol yielding *T. saccharolyticum*, while allowing for ATP generation at 4 moles of ATP per mole of cellobiose.



**Figure 1:** Glycolysis with ethanol production pathway in C. thermocellum: Reactions and cofactors shown in black are present in all cases. The ones in red are replacement candidates by the ones in green. Specifically 1) cellobiose hydrolase (CELLBP) with betaglucosidase (BGL) 2) GTP/GDP with ATP/ADP as cofactors for glucokinase 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, 10) NADH/NAD+ with NADPH/NADP+ as cofactors for alcohol dehydrogenase, 11) H+-translocating ferredoxin:NAD+ oxidoreductase (RNF) with ferredoxin:NAD+ oxidoreductase (NADH-FNOR), and 12) NADH-dependent reduced ferredoxin:NADP+ oxidoreductase (NFN)with ferredoxin:NADP+ oxidoreductase (NADPH-FNOR) .

**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+ |
| RNF\_ATPase | 2 fdxrd + 1 nad+ + 0.25 adp + 0.25 pi <=> 2 fdxox + 1 nadh + 0.25 atp + 0.25 h2o | NADH-FNOR | 2 fdxrd + nad+ <=> 2 fdxox + nadh |
| 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 | NADPH-FNOR | 2 fdxrd + nadp+ <=> 2 fdxox + nadph |

## Assessment of the thermodynamic consistency of the measured metabolite concentrations

In a previous study, the addition of ethanol was shown to inhibit *C. thermocellum* metabolism at the GAPDH reaction [6]. It was suggested that the inhibition was due to regulatory behavior, but an alternative hypothesis is that the inhibition may be due to one or more reactions reaching thermodynamic equilibrium. To test this hypothesis, we re-analyzed the data from Tian et al using the MDF algorithm [1]. The dataset spans intracellular metabolites collected from wild type (WT) *C. thermocellum* growing with and without (control) added ethanol for two replicates each at three different time points. The ethanol concentration is steadily increased to reach a maximum concentration (i.e. at the final time point) of 40 g/L, at which point growth is completely inhibited. The depletion of extracellular cellobiose pool and increase in fermentation products (such as lactate, acetate, ethanol) (see Supplementary table FFF), confirms that cellobiose is converted to ethanol during all three time points for the no-ethanol control but only for the first two time points upon ethanol addition implying thermodynamic feasibility of the ethanol production pathway for at least time points 1 and 2. However, under added ethanol the cellobiose consumption is reduced (time points 1, 2) with no cellobiose consumption beyond time point 3.

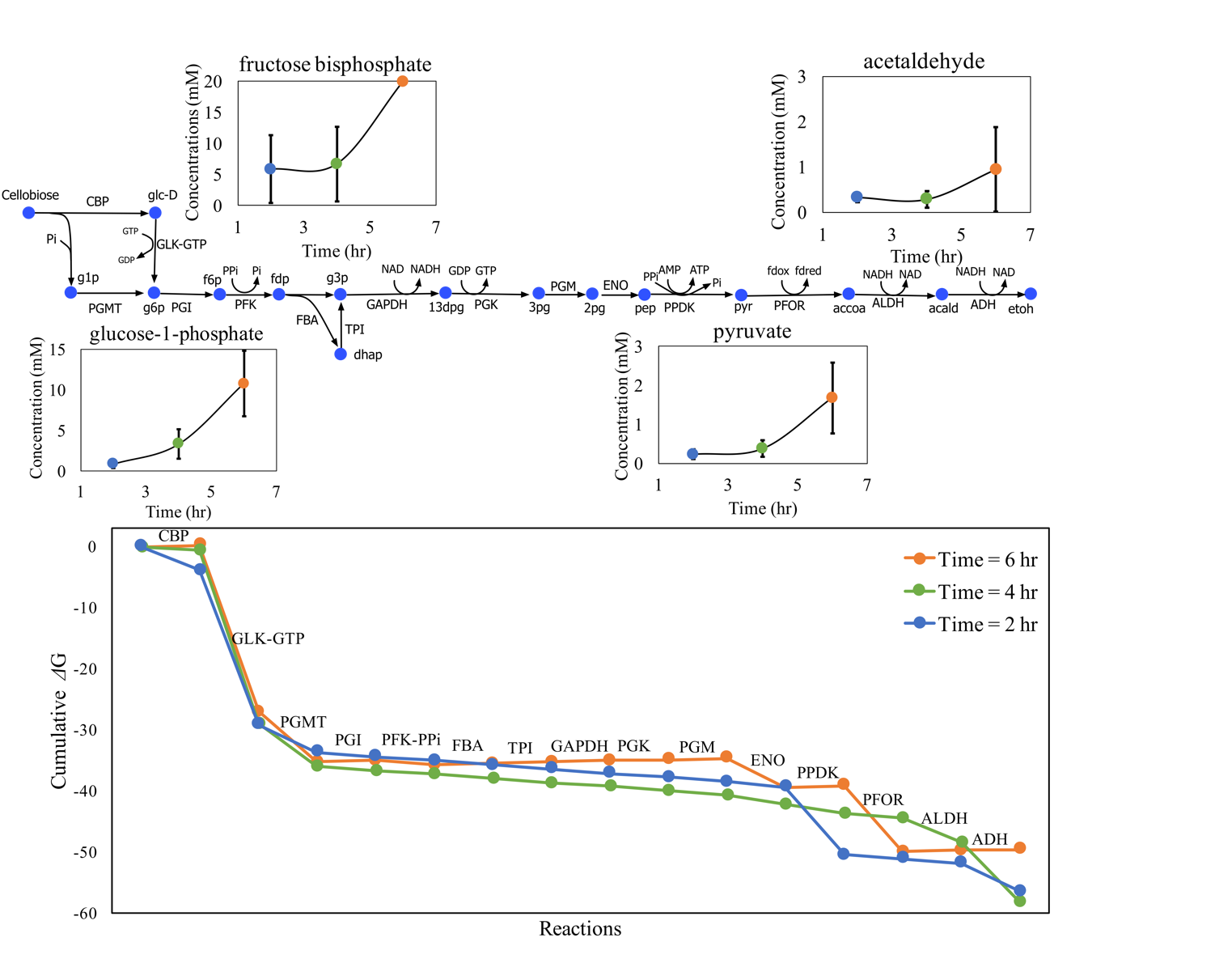
We first explored if the measured metabolite concentrations are thermodynamically feasible (i.e. allow for a positive MDF value) for all 3 time points of the no-ethanol control and for the first 2 time points of the added-ethanol study. In the absence of measurement, metabolite concentrations were allowed to vary within physiological range (1 μM - 0.02 mM), energy cofactor ratios (ATP/ADP, GTP/GDP, and PPi/Pi) were constrained to greater than 10:1 and redox cofactor ratios (NADH/NAD+, NADPH/NADP+, Fd(red)/Fd(ox)) were allowed to vary between 1:100 and 100:1 based on observations in *E. coli* and other clostridia when carbon substrate is not limiting [1, 17-19]. Known concentrations were allowed to vary only within 50% of their measured values (one at a time) allowing liberally for possible measurement error and the pathway MDF (positive MDF implies thermodynamic feasibility) was calculated. The results indicate that the measurements for dhap and fdp lead to negative MDF (with other metabolites varying within physiologically meaningful concentrations [1]) for all time points implying error in quantitation (see Supplementary table GGG). Dhap is difficult to measure by mass spectrometry because it has the same mass to charge (M/z) ratio as glyceraldehyde 3-phosphate (g3p). The two metabolites interconvert by the triosephosphate isomerase (TPI) reaction, potentially allowing for rapid equilibration during quenching. Therefore, it is likely that the measurement error for dhap is higher than 50%. Fdp is measured by liquid chromatography–mass spectrometry (LC-MS) [20], so its lack of thermodynamic feasibility is puzzling. In the no-added-ethanol samples, the fdp levels are around 10 µM, however values would need to be 5-fold higher (at least 50 µM) to ensure thermodynamic feasibility. Although resolving this discrepancy is not the main point of this work, we recognize that this is an important task for future metabolomic studies in *C. thermocellum*. Measured values for dhap and fdp were not considered in the subsequent analyses.

## Glycolysis in *C. thermocellum*

The reduced metabolomic dataset (i.e. ethanol, 3pg, accoa, f6p, g6p, g1p, glc-d, pep, pyr, and mal-l) as described in the previous section were imposed as constraints 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 that the pathway MDF constrained by measured g6p and f6p concentrations (Supplementary table XX) but the overall ethanol production pathway is feasible for all cases except for the final time point upon ethanol addition. For the final time point, the increase in ethanol concentration lowers the driving force of ADH reaction and leads to NADH accumulation. This lowers the driving force of NADH-generating GAPDH reaction due to product (i.e., NADH) accumulation. Thus, GAPDH emerged as the thermodynamic “choke point” which is consistent with experimental observations by Tian et al [6]. The thermodynamic analysis also predicted accumulation of upper glycolysis metabolites especially sugar phosphates (also observed in ethanol stress studies [6, 21]) as a consequence of GAPDH being a thermodynamic bottleneck (Figure 2).

**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 |



**Figure 2:** Glycolysis with ethanol production pathway in wild-type C. thermocellum with PPDK showing the thermodynamic analysis predicted variation in metabolite concentrations of key metabolites across time points. The external ethanol concentration increases along the time points. Sugar phosphates (e.g. g1p) in upper glycolysis are accumulated upon ethanol addition. The bottom panel shows the thermodynamic profile of the ethanol production pathway for the three different time points shown as change in Gibbs free energy of the reactions. The thermodynamic infeasibility for the final time point is a consequence of the positive slope of GAPDH, ALDH and ADH reactions.

## Ethanol pathway modification to relieve thermodynamic barriers

Thermodynamic analysis of WT *C. thermocellum* suggested that the size of cofactor pools (especially NADH) play an important role in the thermodynamic feasibility of ethanol production in *C. thermocellum*. Thus, perturbation in cofactor pools emerged as a primary target for increasing the pathway MDF. To this end, we exhaustively explored the effect of modifying the cofactor association of a single or multiple reactions in *C. thermocellum* on the pathway MDF. Three alternatives to the native reactions (such as ATP linked glucokinase (GLK-ATP), ATP linked phosphoglycerate kinase (PGK-ATP), and malate shunt) and nine heterologous reactions: beta-glucosidase (BGL), ATP-linked phosphofructokinase (ATP-PFK), NADP-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPN), pyruvate decarboxylase (PDC), NADPH-linked ALDH (ALDH-NADPH), NADPH-linked ADH (ADH-NADPH), ferredoxin:NAD+ oxidoreductase (NADH-FNOR), and ferredoxin:NADP+ oxidoreductase (NADPH-FNOR) (see Figure 1 and Table 1) were considered. A change in the reaction’s cofactor preference often implies changes in the corresponding cofactor availability. For example, changing the cofactor preference of the ADH reaction from NADH to NADPH requires that another step in the pathway must regenerate NADPH to ensure cofactor balance. We used the concept of elementary flux modes (EFMs) to formally capture all redox balanced combinations of reactions that allow conversion of one molecule of cellobiose into four molecules of ethanol while distinguishing between reactions that carry out the same conversion but rely on different cofactors. For each one of the identified 336 EFMs (Supplementary Table XX) we calculated the ATP generated, the pathway MDF (at 1M ethanol) and the number of reaction modifications (i.e., cofactor swap) or additions/deletions. Figure 3 depicts the calculated MDF values, ATP generation, and number of modifications of all the EFMs with respect to WT.

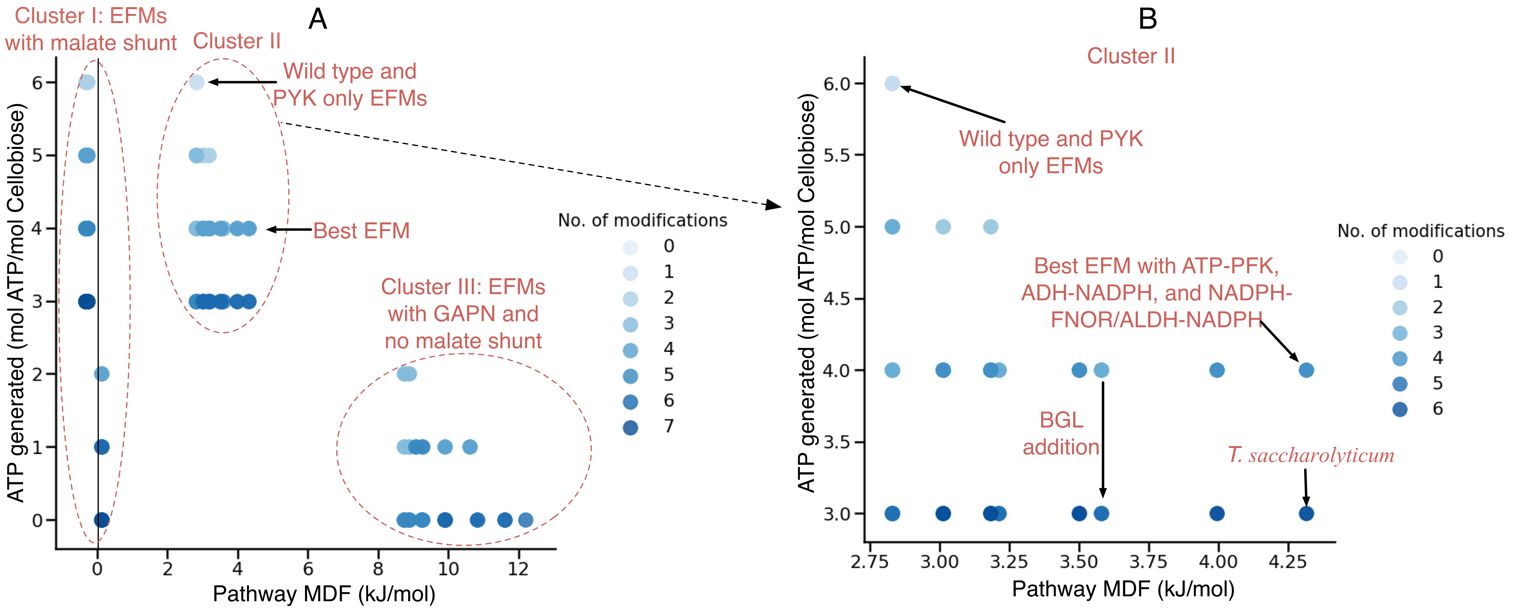
****

Figure 3: A) MDF at 1M ethanol concentration, ATP generated, and number of modifications in all 336 EFMs with three distinct Clusters I: EFMs with malate shunt, II: wild-type EFMs, Best EFMs with high MDF and ATP generation, and III: EFMs with GAPN and no malate shunt. B) Cluster II showing WT C. thermocellum, T. saccharolyticum, and the best EFM with only ATP-PFK, ADH-NADPH and NADPH-FNOR/ALDH-NADPH to support NADPH generation. The EFMs are generated using all possible combinations of cofactor modifications as listed in Table 1. The shade of blue denotes the number of modifications in any EFM from the wild-type EFM.

Three distinct clusters of EFMs emerged upon plotting pathway MDF vs. ATP generated as seen in Figure 3. Cluster I: EFMs with malate shunt and low MDF (<1 kJ/mol), Cluster II: EFMs with intermediate MDF (3-5 kJ/mol) but high ATP generation (>2 mol ATP/ mol cellobiose), and Cluster III: EFMs with GAPN and high MDF (>8 kJ/mol) but low ATP generation (<3 mol ATP/ mol cellobiose).

## Cluster I (malate shunt)

All of the EFMs in Cluster I contain the malate shunt (i.e. phosphoenolpyruvate is converted to pyruvate via oxaloacetate and malate) and are either thermodynamically infeasible (i.e. negative MDF) or have an MDF value slightly greater than zero (i.e., 0.13 kJ/mol) for cases with GAPN (the impact of GAPN is more prominent in Cluster III as discussed in the latter section). *C. thermocellum* does not possess the conventional pyruvate kinase (PYK) enzyme, instead it uses either pyruvate phosphate dikinase (PPDK) or the malate shunt (phosphoenolpyruvate carboxykinase (PEPCK), malate dehydrogenase (MDH), and malic enzyme (ME)) to convert phosphoenolpyruvate to pyruvate (see Figure 1). The EFMs in Cluster I show that PPDK (WT-EFM) is more thermodynamically favorable than its counterpart malate shunt. It is known that malate shunt functions *in vivo* in *C. thermocellum* [16], and it is thought that the purpose of the pathway is to supply reduced NADPH for anabolism. However, the thermodynamic feasibility of PEPCK reaction is limited by intracellular CO2 and oxaloacetate (oaa) concentrations. CO2 concentration is fixed at 10 µM which corresponds to standard atmospheric condition of 400ppm CO2(g), a higher intracellular concentration (> 30 µM for PEPCK feasibility) would allude to supersaturation of CO2 in the media during fermentation as observed by Blunt et al [22]. The sensitivity of the PEPCK reaction to the CO2 concentration may explain why *C. thermocellum* grows much better in the presence of CO2, whether enriched in the atmosphere (typically at 10% v/v) or added as bicarbonate in the growth medium [23]. An alternate driver of PEPCK thermodynamic feasibility is depletion of the oxaloacetate (oaa) pool (< 0.3 µM) due to its spontaneous decarboxylation which also impacts oaa detection in *C. thermocellum* [24]. The challenges in CO2 and oaa measurements can be circumvented indirectly by experimentally measuring other reactants and products in the malate shunt pathway to inform us of *C. thermocellum’s* ethanol production pathway thermodynamics.

## Cluster II

The most highly varied pathway modifications belong to Cluster II. They all retain ATP generation and also significantly increase pathway MDF as shown in Figure 3B. We observe that the presence of pyruvate kinase (PYK) in an EFM does not affect the MDF or the ATP generated by the pathway. On the other hand, betaglucosidase (BGL) addition keeps MDF unchanged but reduces the ATP generation by 1 mol ATP/mol cellobiose (see Figure 3B). The best EFM in Cluster II involves three modifications: PFK-ATP, ADH-NADPH, and ALDH-NADPH/NADPH-FNOR (Supplementary Table XX) with an MDF of 4.3 kJ/mol and 4 moles of generated ATP. The inclusion of ATP-PFK leads to a higher MDF at the expense of ATP generation (the native PFK consumes PPi which is equivalent to 0.5 ATP [24]) while the NADPH linked ADH reaction decouples the impact of ethanol concentrations on NADH associated reactions. ALDH-NADPH or NADPH-FNOR reaction is required to account for NADPH generation and ensure redox balance (Figure 3B). It is interesting to note here that studies have also shown that the cofactor specificity of alcohol dehydrogenase changes to NADPH from NADH in high ethanol yielding strains of *C. thermocellum* [25]. We also observe that EFM-46 (see Supplementary table XX and Figure 3B) corresponds exactly to native glycolysis in *T. saccharolyticum,* a thermophilic ethanologen with high ethanol yields [15] and also has the largest MDF butwith 3 moles of generated ATP in Cluster II based on seven modifications from WT *C. thermocellum* (see Figure 4). Thus, only three genetic interventions (ATP-PFK and NADPH-ADH) to WT *C. thermocellum* are sufficient to maintain high MDF and high ATP theoretical yield without having to fully convert glycolysis to that of *T. saccharolyticum* ethanologen. Of course, thermodynamic feasibility and high ATP yield are necessary but not sufficient conditions for high ethanol titers. Additional bottlenecks may lie with enzyme kinetics and/or pathway regulation.

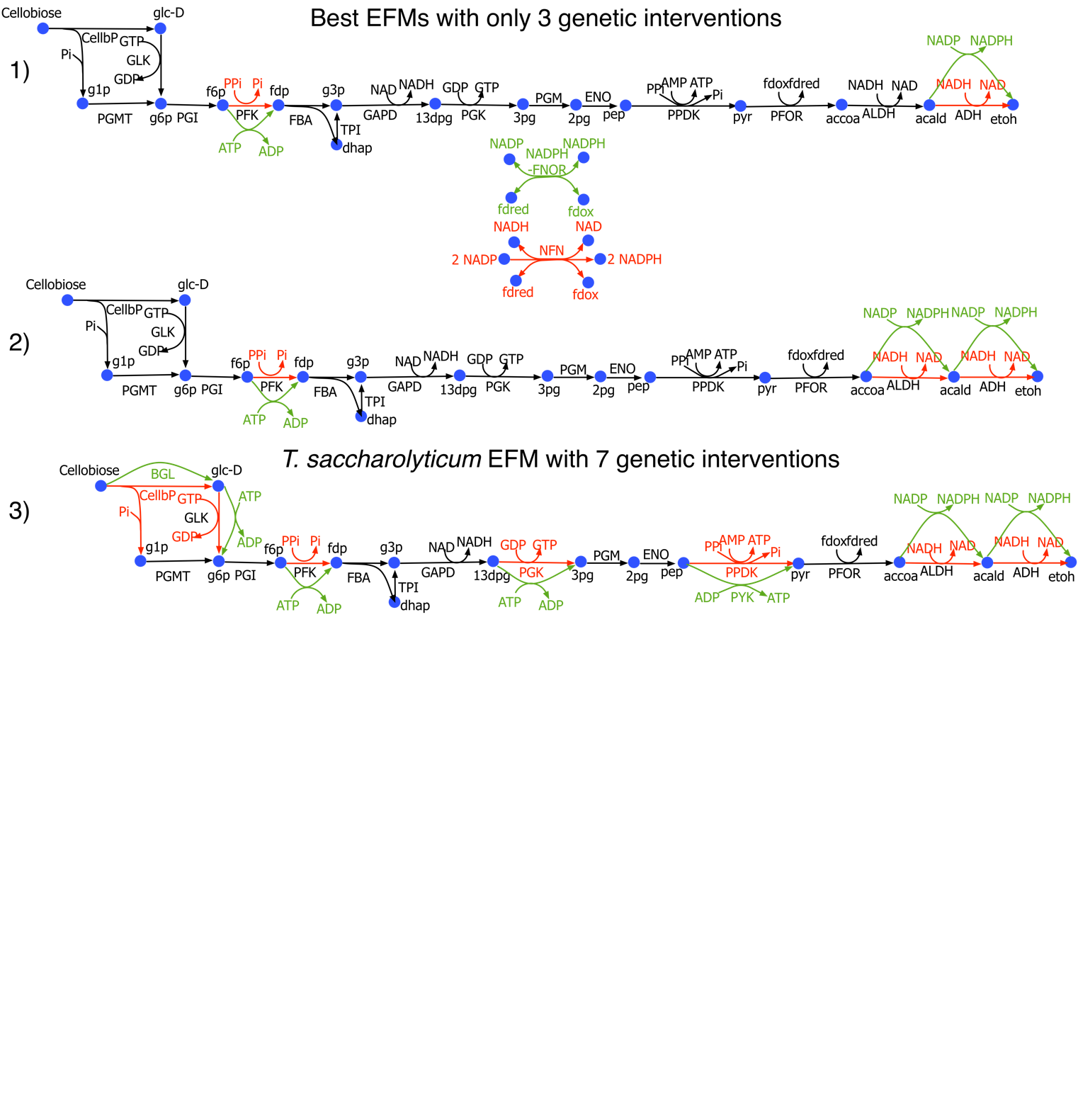


Figure 4: Metabolic pathway map for the best EFMs in Cluster II with only three genetic interventions compared to the EFM which corresponds to T. saccharolyticum: 1) The best EFM with ATP-PFK, ADH-NADPH and NADPH-FNOR, 2) The best EFM with ATP-PFK, ADH-NADPH and ALDH-NADPH. 3) T. saccharolyticum EFM with BGL, ATP- PFK, ATP-PGK, PYK, ALDH-NADPH, and ADH-NADPH. All the EFMs here have an MDF of 4.32 kJ/ mol, but the best EFMs generate 4 mol ATP/mol cellobiose compared to 3 mol ATP/mol cellobiose by T. saccharolyticum EFM.

## Cluster III (GAPN without malate shunt)

All the EFMs in Cluster III involve the GAPN reaction which replaces GAPDH and PGK reactions from wild-type *C. thermocellum*. Previous studies have shown that this intervention can improve ethanol yield in yeast and lysine yield in *Corynebacterium glutamicum* [26, 27]. GAPN mitigates the thermodynamic barrier posed by GAPDH as highlighted in the earlier sections due to rising NADH concentrations by changing the cofactor dependency to NAPDH. However, introduction of GAPN also reduces the ATP generated by the pathway which is necessary for cell growth. The EFM with the highest MDF in Cluster III completely removes NADH association with the ethanol production by relying on PDC and NADPH linked ADH. PDC replaces PFOR and ALDH-NADH thus negating regeneration of NADPH or NADH from reduced-ferredoxin. High GTP and reduced-ferredoxin pools in *C. thermocellum* negatively impacts pathway MDF via PGK and PFOR respectively due to higher product concentration. Thus, inclusion of GAPN and PDC increases pathway MDF. It is also interesting to note 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 higher MDF in engineered strains [28]. In summary, the EFMs in cluster III meet pathway MDF requirements but fall short in terms of ATP generation to support the required growth. Therefore, these interventions could be useful in cases where product formation is decoupled from cell growth [29].

**Conclusions**

This study analyzes wild-type *C. thermocellum* glycolysis using metabolomic datasets to understand the impact of increasing ethanol concentration on pathway thermodynamics. GAPDH is revealed as the major bottleneck at high ethanol concentrations which renders ethanol production thermodynamically infeasible. Plausible pathway modifications were explored and described using the EFM concept. Results revealed the thermodynamic disadvantage of the malate shunt (over PPDK) under physiological CO2 concentrations and the benefit of introducing the GAPN reaction (instead of GAPDH and PGK). GAPN increases the MDF and resolves the GAPDH bottleneck but at the expense of reduced ATP generation at high ethanol concentrations. The study identified the combination of ATP linked PFK and NADPH linked ADH with NADPH liked ALDH or NADPH-FNOR reactions as the best set of genetic interventions which retains ATP generation while maintaining a high MDF similar to the thermophilic ethanologen *T. saccharolyticum*. It must be emphasized that this analysis is based on a partial list of measured pathway metabolites and increasing the number of measured metabolites (particularly cofactors, which participate in many reactions), determining the problems with dhap and fdp quantification, and improving the quantification accuracy (which would allow reducing the error factor below the current value of 50%) will further improve our understanding of the thermodynamic landscape.

It is also important to highlight some of the assumptions and limitations of the MDF analysis carried out in this study. The thermodynamics of cellulose degradation was not considered here due to the heterogeneity of possible products, presence of both intra- and extra-cellular steps and uncertainties on free energy values [2, 3]. Instead the analysis was carried out with intracellular cellobiose as the starting point. Other pathways which produce acetate and amino acids that may interact with the ethanol producing pathway through the cofactor pools were also not considered in the analysis. This was because MFA studies [16, 30] have shown that the ethanol pathway is the major flux carrier. Furthermore, we assumed that the intracellular concentration of ethanol to be the same as the extracellular concentration, since the cytoplasmic membrane is not a significant barrier to ethanol diffusion [31]. However, this assumption may not hold true for other potential biofuel molecules such as butanol warranting intracellular measurements for accurate quantification of MDF. Despite all these simplifications, results obtained demonstrate that MDF analysis is an important first step for assessing thermodynamic feasibility of a pathway and for proposing modifications. Of course, thermodynamic feasibility does not necessarily imply metabolic feasibility as both kinetics and regulation must be sufficient to maintain the required metabolic flows. Follow up investigations of the pathways for their enzyme kinetics and possible regulations are likely to be important [12, 32] but they will still have to operate within the confines defined by thermodynamics.

# Materials and methods

## Metabolite quantification

A single 200 ml culture was grown to an OD600 of 0.1, and then 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 [6, 20, 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 50% during our simulations i.e. the measurements could vary by a factor of 1.5-fold. 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 [20], and thus datasets with 0 energy charge which indicate improper quenching and extraction 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. Metabolites are typically diluted during the quenching and extraction process, to determine the intracellular concentration of metabolites, thus 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 [33]. 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).

## Assessing the thermodynamic feasibility of a pathway

The thermodynamic feasibility of a given pathway is assessed using the max-min driving force formulation [1]. 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 ranges. For performing the MDF analysis, maximum and minimum metabolite concentration bounds were established using the data from Tian et al which 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 time points. Several cofactors were excluded from analysis due to measurement problems. 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. Although Noor et al [1] fixed the ratios of several cofactor pairs, we have relaxed these constraints based on other clostridia when substrate is not limiting [1, 17-19], since the values have not been experimentally determined for *C. thermocellum*. For non-measured metabolites, Noor et al [1] proposed a range of 1 µM to 10 mM, based largely on the work of Bennet et al [17]. Based on our measurements, we decided to keep the default lower concentration at 1 µM, but raise the default upper concentration from 10 mM to 20 mM. Default bounds for each metabolite is described in Supplementary table EEE. The MDF problem is solved using Gurobi Optimizer v6.5.1 solver and Python script modified from the Equilibrator-API Python package [34].

## EFM evaluation

We implement the algorithm to generate k-shortest EFM of a network to systematically evaluate all the EFMs associated with our network [35]. The formulation which largely matches the one of OptStrain [36] 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*, *K* represents the set of previously found solutions and indicates the optimal value of the binary variable in the *kth* solution . The binary variable (constraint 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). Reaction pairs which should not appear in the same EFM can be excluded by adding an exclusivity constraint (i.e., only one reaction from the pair is active in any given solution) on their corresponding binary variables. We finally use integer cuts (constraint 12) to generate all possible EFMs associated with the network. 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. 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:**e1003483.

2. Lynd LR: **The grand challenge of cellulosic biofuels.** *Nat Biotechnol* 2017, **35:**912-915.

3. Lynd LR, Liang XY, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE: **Cellulosic ethanol: status and innovation.** *Curr Opin Biotech* 2017, **45:**202-211.

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

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

6. 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.

7. Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, Yang SH, Land ML, Klingeman DM, Bhandiwad A, Rodriguez M, et al: **Mutant alcohol dehydrogenase leads to improved ethanol tolerance in Clostridium thermocellum.** *P Natl Acad Sci USA* 2011, **108:**13752-13757.

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

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

10. Shao XJ, Raman B, Zhu MJ, Mielenz JR, Brown SD, Guss AM, Lynd LR: **Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of Clostridium thermocellum.** *Appl Microbiol Biot* 2011, **92:**641-652.

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

12. 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.

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:**171.

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

15. Hon S, Olson DG, Holwerda EK, Lanahan AA, Murphy SJL, Maloney MI, Zheng T, Papanek B, Guss AM, Lynd LR: **The ethanol pathway from Thermoanaerobacterium saccharolyticum improves ethanol production in Clostridium thermocellum.** *Metab Eng* 2017, **42:**175-184.

16. 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.

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

18. 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:**D750-753.

19. Meyer CL, Papoutsakis ET: **Increased Levels of Atp and Nadh Are Associated with Increased Solvent Production in Continuous Cultures of Clostridium-Acetobutylicum.** *Appl Microbiol Biotechnol* 1989, **30:**450-459.

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

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. Blunt WA, Gapes DJ, Sparling R, Levin DB, Cicek N: **Quantitative assessment of H2 and CO2 supersaturation during thermophilic cellobiose fermentation with Clostridium thermocellum.** In *2015 ASABE Annual International Meeting*. American Society of Agricultural and Biological Engineers; 2015: 1.

23. Xiong W, Lin P, Magnusson L, Warner L, Liao J, Maness P, Chou K: **Discovery of CO2-fixing one-carbon (C1) metabolism in a cellulose degrading bacterium Clostridium thermocellum.** *Abstr Pap Am Chem S* 2017, **253**.

24. 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:**3000-3008.

25. 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:**2610-2619.

26. 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:**935-943.

27. 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.

28. 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**.

29. Holwerda EK, Thorne PG, Olson DG, Amador-Noguez D, Engle NL, Tschaplinski TJ, van Dijken JP, Lynd LR: **The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading.** *Biotechnol Biofuels* 2014, **7:**155.

30. Xiong W, Lin PP, Magnusson L, Warner L, Liao JC, Maness PC, Chou KJ: **CO2-fixing one-carbon metabolism in a cellulose-degrading bacterium Clostridium thermocellum.** *P Natl Acad Sci USA* 2016, **113:**13180-13185.

31. Shinoda W: **Permeability across lipid membranes.** *Bba-Biomembranes* 2016, **1858:**2254-2265.

32. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R: **Proteomic analysis of Clostridium thermocellum core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression.** *BMC Microbiol* 2012, **12:**214.

33. 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**.

34. Noor E, Haraldsdottir HS, Milo R, Fleming RM: **Consistent estimation of Gibbs energy using component contributions.** *PLoS Comput Biol* 2013, **9:**e1003098.

35. 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:**3158-3165.

36. Pharkya P, Burgard AP, Maranas CD: **OptStrain: A computational framework for redesign of microbial production systems.** *Genome Research* 2004, **14:**2367-2376.