# Materials and methods

## Metabolite quantification

Dataset 2 from Tian et al. 2017 (Tian et al., 2017) was used. Briefly, 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 3 timepoints (T=2.0h, 3.8h and 5.9h), each culture was sampled twice for intracellular metabolites using previously described protocols (Olson et al., 2016; Rabinowitz and Kimball, 2007; Tian et al., 2017), 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 4 measurements of our standard curves, our quantification error for

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 (Rabinowitz and Kimball, 2007), and were excluded from analysis.

(Equation 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.

(Equation 2)

The intracellular volume is assumed to be 3.9 µL for 1 ml of a culture at an OD600 density of 1 (Volkmer and Heinemann, 2011). 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.

## Thermodynamic model constraints

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. (Noor et al., 2014) proposed a range of 1 µM to 10 mM, based on largely on the work of Bennet et al. (Bennett et al., 2009). 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 cofactors were set to allow ratios between 1:100 and 100:1 for the pairs (ATP/ADP, ATP/AMP, GTP/GDP, NADH/NAD+, NADPH/NADP+, Fd(red)/Fd(ox)). Default bounds for each metabolite is described in Supplementary table EEE.

## Stoichiometric model

Based on XXX. Previous stoichiometric models of C. thermocellum have ignored cofactor balance. To reintroduce some of that complexity back into our model, we…

## EFM generation

Constraints for EFM generation:

1. 1 cellobiose is consumed and 4 ethanol are produced
2. All reactions are in the forward direction except NDK reaction and H2O exchange reaction
3. Net ATP production is the sum of ATPase1 and ATPase2. 1 ATP is subtracted from this number to account for cellobiose transport.

# Results and discussion

## Description of metabolite dataset

Dataset 2 from Tian et al. 2017 (Tian et al., 2017) was used. This dataset represents intracellular metabolites collected from WT *C. thermocellum* growing with and without the presence of added ethanol. The maximum concentration of added ethanol is 40 g/L, which is thought to be a minimum titer for commercial viability (Dien et al., 2003).

## Thermodynamic consistency of dataset

Based on measurements of extracellular metabolites (Supplementary table FFF) we know that cellobiose was converted to ethanol during all 3 timepoints for the no-ethanol control, which implies that all reactions must have a positive MDF. For the samples with added ethanol, we cannot directly measure the ethanol flux, since the added ethanol is much larger than the produced ethanol, but based on the consumption of cellobiose and the presence of other fermentation products (formate and acetate), we suspect that there was flux from cellobiose to ethanol at the first two timepoints (and thus MDF should be positive). For the 3rd timepoint, flux was close to zero. The expected MDF values are summarized in Table GGG.

**Table GGG. Expected MDF values for ethanol addition experiment**

|  |  |  |
| --- | --- | --- |
| Timepoint | Added ethanol | Control |
| T0 | + | + |
| T1 | + | + |
| T2 | ≤ 0 | + |

We attempted to fit all of the measured metabolite concentrations (excluding cofactors) to our thermodynamic model. For each measured metabolite, we included an uncertainty factor of 2 (see materials and methods for description). This resulted in negative MDF values for both the added ethanol and control cultures at all timepoints, indicating that the dataset is not thermodynamically consistent. To identify which metabolites were causing the inconsistency, we tested them one-by-one (Table HHH).

**Table HHH. MDF values for each metabolite tested one-by-one from each fermentation condition.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **ID** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|  | **graphName** | **A1\_1 et=5.0** | **A1\_2 et=5.0** | **A2\_1 et=20.0** | **A2\_2 et=20.0** | **A3\_1 et=40.0** | **A3\_2 et=40.0** | **B1\_1 et=0.2** | **B1\_2 et=0.2** | **B2\_1 et=0.47** | **B2\_2 et=0.47** | **B3\_1 et=0.58** | **B3\_2 et=0.58** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **C00010** | **coa** | 2.3 | 2.2 | 2.4 | 2.5 | 2.5 | 2.5 | 2.2 | 2.2 | 2.4 | 2.4 | 2.2 | 2.1 |
| **C00469** | **etoh** | 2.7 | 2.7 | 2.3 | 2.3 | 2.1 | 2.1 | 3.5 | 3.5 | 3.4 | 3.4 | 3.4 | 3.4 |
| **C00197** | **3pg** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
| **C00024** | **accoa** | 0.6 | 0.3 | -1.1 | 2.1 | -1.2 | -1.2 | -0.5 | -0.2 | -0.8 | -1.2 | -0.1 | 0.1 |
| **C00111** | **dhap** | -5.5 | -4.7 | -2.3 | -2.0 | -2.0 | -1.7 | -4.8 | -4.3 | -4.6 | -5.8 | -3.6 | -5.2 |
| **C00085** | **f6p** | 1.8 | 1.8 | 1.9 | 1.9 | 2.1 | 2.0 | 1.9 | 1.9 | 1.9 | 1.9 | 1.8 | 1.1 |
| **C00354** | **fdp** | -1.5 | -1.2 | -0.3 | -0.4 | -0.3 | 0.0 | -0.9 | -0.8 | -1.1 | -1.0 | -1.0 | -0.9 |
| **C00092** | **g6p** | 1.5 | 1.4 | 1.5 | 1.5 | 1.5 | 1.6 | 1.5 | 1.5 | 1.6 | 1.5 | 1.7 | 1.8 |
| **C00103** | **g1p** | 1.6 | 1.8 | 2.1 | 2.1 | 2.1 | 2.1 | 1.9 | 1.9 | 2.0 | 2.1 | 2.1 | 2.1 |
| **C00031** | **glc-d** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
| **C00074** | **pep** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
| **C00022** | **pyr** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
| **C00149** | **mal-l** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |

From this, we can see that there are potential problems with 3 metabolites: acetyl-coA (accoa), dihydroxyacetonephosphate (dhap), and fructose-1,6-bisphosphate (fdp).

## FDP and DHAP levels

FDP and DHAP are the two metabolites that cause the biggest problems with thermodynamic consistency of the *C. thermocellum* metabolic pathway. The set of reactions that allow for substrate-level phosphorylation (SLP) in EMP glycolysis (FBA, TPI, GAPDH and PGK) are known to be a thermodynamic bottleneck (Noor et al., 2014; Stephanopoulos et al., 1998). In *E. coli*, this pathway is feasible due to the high (>10 mM) intracellular concentration of FDP (Bennett et al., 2009), however in *C. thermocellum*, the intracellular concentration of FDP is 2-3 orders of magnitude lower (10-100 µM). At 10 µM, none of the datasets from the ethanol addition experiment are thermodynamically feasible. At 100 µM, however, 11 of 12 datasets are consistent with the MDF predictions from Table GGG, 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.

In these bottleneck reactions (FBA, TPI, GAPDH and PGK), there are several cofactor ratios that contribute to the thermodynamic feasibility: NAD/NADH and ATP/ADP. In the Noor et al. analysis (Noor et al., 2014), both of these ratios were fixed at a value of 10. In our analysis, we have provided additional degrees of freedom by allowing these cofactor ratios to vary over a wide range (100 to 0.01 for the NAD/NADH ratio and 200 to 10 for the ATP/ADP ratio). For the *C. thermocellum* datasets that are thermodynamically feasible with FDP at 100 uM, the NAD/NADH ratio is typically close to the upper end of the range (i.e. 100) and the ATP/ADP ratio is close to the lower end of the range (i.e. 10). Thus, our model allows lower FDP levels by increasing the NAD/NADH ratio. Since the NAD/NADH ratio decreases as ethanol titer increases, this may make *C. thermocellum* more susceptible to ethanol inhibition than other organisms such as *E. coli*.

In our model, DHAP levels need to be about 1 mM for thermodynamic feasibility, which is 20-fold higher than the range of 5-50 µM that we measured. 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 than for FDP.

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.

## Conclusions for ethanol addition dataset analysis

1. *What do we know about causes of ethanol inhibition?* Here we applied a thermodynamic framework to the measured metabolite concentrations to “upgrade” the data. In general, we find a metabolic bottleneck distributed among the SLP reactions, consistent with our previous analysis of the data (Tian et al., 2017). In addition, we find the metabolic bottleneck is also distributed between the ADH and ALDH reactions. This highlights the role of the NAD/NADH ratio in the thermodynamic feasibility of the cellobiose to ethanol pathway in *C. thermocellum*.
2. Low levels of FDP constrain the feasible range of the NAD/NADH ratio.
3. Traditionally, glycolysis is regulated by hexokinase (aka glucokinase), PFK and PYK. In C. thermocellum, the only reaction with a large negative delta G value is glucokinase, so this may be the main regulator of glycolysis.
4. In the next section we investigate possible metabolic engineering interventions to improve ethanol titer in *C. thermocellum*.

## Analysis of thermodynamic landscape of *C. thermocellum* and possible metabolic engineering interventions

One goal of this work was to provide guidance for metabolic engineering efforts. To do this, we decided to look at the thermodynamic landscape (i.e. the range of metabolite concentrations that are thermodynamically feasible for different pathways that *C. thermocellum* can use to convert cellobiose to ethanol). In addition to the native reactions, we included several additional reactions that we are considering incorporating into *C. thermocellum*, including: pyruvate decarboxylase, ATP-linked phosphofructokinase, beta-glucosidase, NADPH-linked ALDH, NADPH-linked ADH, a non-energy-conserving FNOR reaction (the native *C. thermocellum* FNOR reaction is RNF, which conserves energy by pumping ions across the membrane).

A metabolic 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 generated elementary flux modes (EFMs) for all combinations of reactions that allow conversion of one molecule of cellobiose into 4 molecules of ethanol. For each of the resulting 336 EFMs, we calculated the ATP yield and the MDF. To identify how many changes had occurred in a strain, we also measured the Hamming Distance between different flux vectors.

This allowed us to answer the following questions:

1. Which metabolic intervention has the biggest impact on MDF?
2. Which metabolic interventions allow MDF to be increased without affecting ATP yield?
3. What is the simplest metabolic engineering intervention (i.e. the one with the smallest Hamming Distance from the WT strain) that has the biggest impact on MDF?

## Additional text that I don’t have a place for yet

Thus, two possible approaches to increase ethanol titer are to increase FDP levels (possibly by changing the PPi-PFK reaction to ATP-PFK) or to break the link between NADH-producing reactions (i.e. GapDH) and NADH-consuming reactions (i.e. ADH and ALDH).