**Replies to reviewers’ comments**

**Reviewer 1**  
  
The study combines theoretical metabolic network analysis with thermodynamics and metabolomics to predict flux limiting reactions in an anaerobic bacterial system. The goal of the approach is to identify rational intervention targets to increase flux from cellobiose to ethanol. The study extends previous experimental and theoretical work by the Ctherm group which previously considered the thermodynamics of the pyruvate to ethanol portion of glycolysis based on metabolite levels.

The study is interesting and combines a range of expertise to tackle a complex problem. The computational approach is useful and will be of interest to a wide range of quantitative biologists and engineers. The material is publishable but only after some substantial adjustments are made.

Major:

The current work does not account for synthesis of byproducts acetate, formate and lactate etc. According to supp file 5, the cultures produce more carbon moles of these byproducts than ethanol so leaving them out of the analysis introduces a very large source of error, additionally the model doesn’t consider biomass although this would be a small flux compared to the byproducts. The authors make an argument (line416+) they are not interested in the byproducts because they aren’t important with high ethanol selectivity strains (Olson, 2015, 2017). This isn’t a good argument because the current study uses metabolite concentrations from a strain that makes byproducts and therefore the data should not be projected onto strains that behave differently. This really needs to be addressed in a more satisfactory manner.

<Answer>

Byproduct synthesis and growth might facilitate the problems with the infeasible 3pg concentrations calculated here.

<Answer>

The authors don’t fix the cofactor pools for simulations and instead allow them to fluctuate within a preset range. An approach that is reasonable when lacking data, but it is unclear what data is lacking. Have the authors compared the predicted cofactor levels to the cofactor data reported in Tian 2017? Looking at the NADH and NAD+ values in the simulation data (supp fig 2), the ratio is NADH/NAD+ = 1/100, this is two orders of magnitude smaller than the experimental values in Tian 2017 (~1/1 ratio); the simulation for NADPH/NADP = 1/25 while the experimental ratio reported in Tian 2017 data is 15/1, again two orders of magnitude different. There seems to be some major issues with the analysis aligning with experimental data or there is confusion as to what experimental data is relevant.

<Answer>

Line 126: it is stated NADH was not observed, however in Tian 2017b, figure 2 there is NADH/NAD+ and NADPH/NAD+ data. It is not clear what data was or was not used from the previous work leading to confusion.

<Answer>

Minor:

Does the model distinguish between CO2 and bicarbonate and their equilibrium? Do the carboxylase enzymes all use CO2 or do some use bicarbonate?

<Answer>

Line 343 does the model account for the spontaneous decarboxylation of OAA, how does this rate compare to the enzyme catalyzed reactions?

<Answer>

Fig 2 only shows 4 of the 8 considered metabolites, adding the other profiles would be useful

<Answer>

Table 1: what is the error of the metabolite measurements, why aren’t all three time points listed to facilitate communication?

<Answer>

Table 2: is this the EFM model? The caption calls it the ‘list of possible reactions in any given EFM’. The unbalanced metabolites should be listed in the table caption.

<Answer>

Table 4: the values need units

<Answer>

The H2 levels in the experiment could have a large effect on the redox cofactor equilibria, was the H2 measured or is there an estimate of the range?

<Answer>

Line 216 vs 324: is this supposed to be positive or negative? MDF (i.e., - 0.13 kJ/mol vs 0.13 kj/mol)

<Answer>

EFMs play a large role in the document, yet they are only introduced in the results section. I would recommend introducing the concept in the introduction and providing a brief summary of earlier work which is relevant. The current work should be framed in terms of some of the earlier EFM thermodynamics and metabolic engineering work: [https://www.sciencedirect.com/science/article/pii/S1096717606000103](https://nam01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.sciencedirect.com%2Fscience%2Farticle%2Fpii%2FS1096717606000103&data=02%7C01%7Ccdm8%40psu.edu%7C43ce4adf338b45f1cc4b08d6c04b0f72%7C7cf48d453ddb4389a9c1c115526eb52e%7C0%7C1%7C636907825949193046&sdata=4ZSzL%2FO5LGlxnVnqLfF5T0AI%2Bb%2B86qxBR1OjXk%2BThMk%3D&reserved=0)

[https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171440](https://nam01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fjournals.plos.org%2Fplosone%2Farticle%3Fid%3D10.1371%2Fjournal.pone.0171440&data=02%7C01%7Ccdm8%40psu.edu%7C43ce4adf338b45f1cc4b08d6c04b0f72%7C7cf48d453ddb4389a9c1c115526eb52e%7C0%7C1%7C636907825949203054&sdata=Pv9MraUKUAWn05e2GafHze2Lq1x29DfkdnZsvypLBP8%3D&reserved=0)

[https://link.springer.com/article/10.1007/s00253-008-1770-1](https://nam01.safelinks.protection.outlook.com/?url=https%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs00253-008-1770-1&data=02%7C01%7Ccdm8%40psu.edu%7C43ce4adf338b45f1cc4b08d6c04b0f72%7C7cf48d453ddb4389a9c1c115526eb52e%7C0%7C1%7C636907825949203054&sdata=3BfE0U7tmHQfo%2B03yAiX7FgQYBp619GhlKp4x0b73Qo%3D&reserved=0)

<Answer>

Line 211 references citation [17] which isn’t consistent with the name and date format of the other references.

<Answer>

Line 333, 335 mu is used on one line and u on the other, double check consistent use of ‘micro’ designator

<Answer>

Line 62: missing the word ‘more’

<Answer>

Line 191: missing an ‘s’ on reaction

<Answer>

-**Reviewer 2**  
  
Overall:

The authors performed a thermodynamic analysis of ethanol production in C. thermocellum with the goal of understanding why production is inhibited at high ethanol concentrations. The analysis is very high-quality and leads to meaningful metabolic explanations. The authors could likely go further in the discussion/conclusions section and address whether their findings change the industrial potential of C. thermocellum for ethanol production.

Questions/Points:

Table 3: Any issues with replacing NAD(H) with NADP(H)? Should cofactor stability be considered?

<Answer>

Ethanol toxicity has been thought to occur due to cell membrane leaking and possibly enzyme inhibition or unfolding. How do you know these factors are not involved in this case with C. thermocellum?

<Answer>

After learning that *T. saccharolyticum* has more favorable metabolism for ethanol production, which organism should be engineered? Shouldn’t the favorable characteristics of *C. thermocellum* be applied to *T. saccharolyticum* and not vice-versa?

<Answer>

It is somewhat unclear what conclusions are being made, and I feel there is plenty of room to establish metabolic and organismal guidelines for ethanol production based on the thermodynamic analysis. Some example questions follow…. Metabolic interventions for C. thermocellum were found, but can these be implemented? If so, what are the expected improvements in production? Is the metabolic engineering strategy too difficult for this organism? Should continuous ethanol stripping during fermentation be used over metabolic engineering? Or, does this analysis suggest C. thermocellum is not the best candidate for ethanol production? Should candidate organisms for ethanol production have a set of core metabolic characteristics to be considered for further engineering? If so, what are they? Can libraries of organisms (especially thermophiles) be scanned for these characteristics?

<Answer>

-**Reviewer 3**

**(Please note that this review was sent directly to me and is not in the system)**

*Clostridium thermocellum* does show arguably the best cellulose degradation capabilities for bioethanol production from cellulosic material using consolidated bioprocessing approaches. While mutant strains have been successfully designed to produce ethanol to near theoretical yields per hexose, its capacity to produce high titers of ethanol is very limited. Based on metabolite concentration measurements and thermodynamic modelling, the present manuscript provides a plausible explanation for this observation, and suggests possible gene replacements that would alleviate thermodynamic pathway bottlenecks at high ethanol titers. From an evolutionary point of view, this research permits one to appreciate how this organism has evolve to maximize energy conservation per mole substrate rather than maintain biological energy pools through fluxing high amounts of substrates to end-products like most better-known fermenting organisms.

<Answer>

This is a very interesting set of findings with significant biotechnological applications for enhancing biofuel production from consolidated bioprocessing. There are, however, several improvements with respect to the clarity of the data presentation that can be suggested along with a few questions about the way some of the data are presented.

<Answer>

Major comments

1. Figure 2 legend. The positive slope of PFK, FBA, GAPDH, ALDH, and ADH reactions are not really visible within any part of this figure. A table of individual delta G values in the supplementals that you could refer the reader to would be helpful in order to appreciate these important findings.

<Answer>

Minor comments

Line 62 – “…is **more** thermodynamically favorable…”

<Answer>

Supplemental file 1 – indicating Units for amounts would be helpful to the readers?

<Answer>

Line 271 – Looks like an “is” might be missing between “MDF” and “constrained”

<Answer>

Line 405 – “NADPH liked ALDH” should be “linked”

<Answer>

Please look at figures 1 and 4 carefully. Some of the arrows are going in the wrong direction. For example, you are indicating in the green arrows that the conversion of acetaldehyde to ethanol used NADP to generate NADPH and you have 1,3pg +ATP yielding 3pg+ATP. I may have missed some other errors

<Answer>

**Major comments:**

(1) In figure 2, I donk think using FVA to evaluate model prediction is a wise choice, especially for secreted products like ethanol. It would be better to do the evaluation using FBA and comparing that with kinetic model prediction and experimental data.

[[1](#_ENREF_1)][[1](#_ENREF_1)] <Answer>

(2) In the section "Effect of nitrogen limitation on model predicted phenotype", it is concluded that "model recapitulated experimental observation that reduction in nitrogen availability reduces the fermentation pathways and reroutes additional flux towards amino acid production". The reference cited is an unpublished work which doesn't verify the finding.

<Answer>

(3) In the section "Effect of ethanol stress on wild-type C. thermocellum", Inorganic diphosphatase (PPA) was not mentioned in ref. 51. Please provide the rationale behind upregulating it by 50%.

[[2](#_ENREF_2)][[2](#_ENREF_2)][[3](#_ENREF_3)][[2](#_ENREF_2)] <Answer>

(4) It makes sense that upregulation of PPA will lead to low PFK activity but the 2.1-fold increase of ammonium owing to low PFK activity might not be correct because the Km value of ammonium in PFK is low (10-4M). It could be the downregulation of ACLS, which is the first reaction in branched-chain amino acid biosynthesis pathway, decreases the flux towards amino acid pathway which will eventually lead to accumulation of ammonium. Please comment on this.

[[3](#_ENREF_3)] <Answer>

(5) The model utility toward biofuel production and strain design strategy should be demonstrated and discussed. The relevant applications and future directions can be also highlighted.

[[4](#_ENREF_4)]. [[5](#_ENREF_5)][[6](#_ENREF_6)]

**Reviewer #2**: The authors present an updated genome-scale flux model and a new kinetic model for Clostridium thermocellum metabolism.  C. thermocellum is of significant interest to cellulosic biofuels, and the approach in this manuscript represents a step forward in metabolic modeling.  Overall, an improved flux model was generated, and the kinetic model showed the ability to offer responsiveness to nitrogen limitation and ethanol stress that cannot be done with flux models.  However, there is an opportunity that has been missed in the current manuscript.  That is how to use data and experimental design for building a good kinetic model.  Here, additional experiments could be suggested that will lead to a more accurate kinetic model in future generations.

Kinetic model simulations direct our attention towards the incomplete metabolic knowledge involving specific pathways, which can be resolved using 13C-MFA, transcriptomic and proteomic studies for wild-type and mutant strains to accurately capture the pathway activity under various conditions. We have updated the discussion section describing clearly the kinetic models insights into *C. thermocellum* metabolism and the relevant experimental suggestions to verify these discoveries (see lines 1-5, lines 10-21 of page 19).

**Major Comments:**

Among the major updates leading to the generation of iCth446 is a significantly lower GAM value.  The authors note the new value used is in-line with other clostridia models.  However, how do this authors know this value is correct?  Could (and should) this value be determined experimentally?  It is common for models to infer biomass and maintenance energy descriptions from other models, but what if those original models are not correct?

We agree with the reviewer on the effect of GAM value on model’s predictions. This value can be determined experimentally as done for several organisms such as *E.coli* [[7](#_ENREF_7)]. However, due to lack of experimental measurements for non-standard organism such as *C. thermocellum* we rely on information from phylogenetically close organisms or models. The GAM assumption should be validated with model predictions under genetic perturbations as highlighted in the manuscript by the comparisons of growth rate predictions by *iCth446* vs existingGSM models under *Δack* conditions [[8](#_ENREF_8)]. In order to address the reviewer’s concern, this is highlighted in the revised manuscript (see lines 1-8 of page 6).

Figure 2.  In the top of Fig. 2, the GSM seems to out-perform the kinetic model.  Yet, the authors claim the kinetic model is ultimately superior due to the ability to incorporate regulatory elements.  This should be addressed in discussion because it appears that the kinetic model is not yet fully representative of metabolism.  However, this is OK because the kinetic approach still represents a significant step forward in modeling C. thermocellum metabolism.

The reviewer is correct that the kinetic model accuracy depends on the range of covered pathways and scope of parameterization datasets. Overall, kinetic models have the potential to substantially outperform stoichiometry-based predictions when parameterized under similar conditions with well-defined metabolic descriptions (such as cofactors and regulations). But kinetic model predictions exhibit inconsistency when queried under significantly different conditions or due to incomplete metabolic descriptions. In general, we observe that the k-ctherm118 model predictions for 10 out of 19 mutant datasets lie within 20% of the experimental value. The error in predictions of the remaining mutants can be explained by secondary mutations which were unaccounted for during model simulations. Error in prediction is also a result of incomplete understanding of *C. thermocellum* metabolism as revealed by robustness analysis. GSM model predictions on the other hand are broad ranges even when constrained with several experimentally measured metabolites. GSM models also do not capture regulatory events, thus limiting their usefulness compared to kinetic models. In order to address the reviewer’s concern, this is now highlighted in the revised manuscript (see line 2-9 of page 20).

Figure 2.  The constraints used to generate this figure would be helpful.  These could be added as a table in the supplementary appendix.

We have added the complete set of fermentation data for all the metabolites along with their associated standard deviations in the revised supplementary file 3 (see *fermentation data* worksheet).

Page 10, line 14.  The reference to mutants in Figure 2, "…(i.e., mutants 1, 4-6, 12, 15, 16, 18 and 19)…" is not clear.  Do these mutants correspond to the gene numbers in Figure 2?

We thank the reviewer for pointing this out. We have updated the Figure 2 to reflect that the numbers correspond to mutants and not genes.

Page 10, line 18.  There is mention to "secondary mutations which are unaccounted for in the various strains (see Table 3)…" Table 3 does not mention secondary mutations.  I do understand the authors are referring to the errors observed in kinetic parameters, but there should be clearer way to make this point.

We refer to the error in the experimental data which is being used to test the stoichiometric model and parameterize the kinetic model. The experimental data for different strains with the same genotype is pooled together during model testing which leads to wide confidence ranges in the training datasets (i.e., an average 50% error in measurement). This error can be explained by the fact that secondary mutations are unaccounted for in the genotype information of several strains which can be resolved by analyzing the sequences of all the pooled strains to ensure that all mutations are accounted for. This has been clarified in the revised manuscript (see line 1-2 of page 10).

The formation and training/testing of the k-ctherm118 kinetic model led to some very interesting insights in my opinion.  While significant error was observed with the kinetic model, this is very valuable because it points to an incomplete understanding of central metabolism. This is something not realized from flux modeling, and it is only apparent when several mutant strains are available for testing.  Can the model training/testing procedure lead to the identification of additional experiments that are need for model validation?  I believe there is an opportunity here to demonstrate the experimental/computational cycle that must occur to build a good kinetic model.

We agree with the reviewer on application of k-ctherm118 to suggest additional metabolic experiments. For example, the robustness analysis has revealed the secondary activity of Ketol acid-reductoisomerase which can be tested experimentally. Identifying flux split ratios at nodes is essential for kinetic model parametrization as observed for the case of error in prediction of kinetic parameters associated with the PP pathway due to lack of accurate flux information in training datasets of k-ctherm118. This can be experimentally elucidated with the aid of 13C-metabolic flux analysis (13C-MFA) data [44]. Transcriptomic and proteomic data for various strains are also essential to determine the perturbed enzyme levels for accurate model parametrization as well as predictions. Thus, kinetic model simulations direct our attention towards the incomplete metabolic knowledge involving specific pathways which can be resolved using 13C-MFA, transcriptomic and proteomic studies for wild-type and mutant strains to accurately capture the pathway activity under various conditions. This is discussed in detail in the discussion section of the revised manuscript (see line 10-21 of page 19).

It seems to me that metabolomics data are needed for full validation of a kinetic model.  Can the authors comment on this in the discussion?  What are the ideal sets of experiments and measurements that must be obtained to develop a good predictive kinetic model?

Kinetic model simulations require complete metabolic knowledge involving specific pathways which can be addressed by using 13C-MFA, transcriptomic and proteomic studies for wild-type and mutant strains to accurately capture the pathway activity under various conditions. In an earlier study [[9](#_ENREF_9)], we observed that incorporation of accurate metabolite concentrations led to accurate estimation of kinetic parameters. With the ever-increasing metabolomic databases (e.g., MetaboLights [[10](#_ENREF_10)]), integration of such information will improve the quality of model parameterization. Other omics platform such as transcriptomic and proteomic are also necessary to recapitulate the changes in enzyme levels (i.e., vmax) in response to genetic and/or environment perturbations [[9](#_ENREF_9)]. These are discussed in detail in the revised manuscript (see line 1-5 of page 19).

The manuscript is critical of the iSR432 model in several places, including page 10, line 8 with the sentence, "The comparison revealed that as expected iSR432 significantly under-predicts fermentation products…"  Instead of this approach, it is recommended that the authors acknowledge iSR432 as an important foundation upon which iCth446 was built.  Then, go on to explain the improved capabilities iCth446.  This can be done without pointing out shortcomings of iSR432.

We appreciate the efforts of Roberts *et al*. [[11](#_ENREF_11)]in building the first stoichiometric model of *C. thermocellum* and which served as an essential foundation of our work. However, it is also essential to highlight the significance of the various metabolic updates to *iSR432* and the metabolic impact of these changes on model predictions. We have updated the manuscript to highlight our acknowledgment on this model (see line 8-9 of page 26).

**Suggested Edits:**

Page 3 (of pdf file), line 3. Change "gram" to "Gram"

Page 3, line 20.  I suggest starting this sentence. "The k-ctherm118 model captures…"

*Page 5, line 8.  Change "gram" to "Gram"*

*Page 8, line 13. Change "models" to "model's"*

We thank the reviewer for pointing these out. We have updated these corrections in the manuscript.

We thank the reviewer for their detailed comments that led to significant improvements in the manuscript.

1. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.** *Metab Eng* 2003, **5:**264-276.

2. 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.

3. 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.** *Proc Natl Acad Sci U S A* 2016, **113:**13180-13185.

4. Lo J, Olson DG, Murphy SJ, Tian L, Hon S, Lanahan A, Guss AM, Lynd LR: **Engineering electron metabolism to increase ethanol production in Clostridium thermocellum.** *Metab Eng* 2017, **39:**71-79.

5. Chowdhury A, Zomorrodi AR, Maranas CD: **k-OptForce: integrating kinetics with flux balance analysis for strain design.** *PLoS computational biology* 2014, **10:**e1003487.

6. Flowers D, Thompson RA, Birdwell D, Wang T, Trinh CT: **SMET: systematic multiple enzyme targeting - a method to rationally design optimal strains for target chemical overproduction.** *Biotechnol J* 2013, **8:**605-618.

7. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol Syst Biol* 2007, **3:**121.

8. Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM: **Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in Clostridium thermocellum.** *Biotechnol Biofuels* 2015, **8:**20.

9. Khodayari A, Maranas CD: **A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains.** *Nat Commun* 2016, **7:**13806.

10. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, et al: **MetaboLights-an open-access general-purpose repository for metabolomics studies and associated meta-data.** *Nucleic Acids Research* 2013, **41:**D781-D786.

11. Roberts SB, Gowen CM, Brooks JP, Fong SS: **Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production.** *BMC Syst Biol* 2010, **4:**31.