molecular systems biology

Proteome-wide systems analysis of a cellulosic biofuel producing microbe

Authors: Andrew Tolonen, Wilhelm Haas , Amanda Chilaka , Dr. John Aach , Steve Gygi , George Church

Corresponding authors: Andrew Tolonen (Harvard Medical School) and Wilhelm Haas (Harvard Medical School)

Review timeline:

Submission date:	09 September 2010
Editorial Decision:	08 October 2010
Revision received:	03 November 2010
Acceptance letter :	30 November 2010

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 08 October 2010

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. However, they raise substantial concerns on the work, which should be addressed in a major revision of the present work.

While reviewer #2 (expertise: systems biology, proteomics) and reviewer #3 (expertise: cellulose degradation) are rather supportive, reviewer #1 (expertise: systems biology, metabolic engineering) is more reserved. One of the major concerns expressed by reviewer #1 (point 5) refers to the current lack of evidence demonstrating that the systems-level insights (summarized in Fig 7) derived from the present analysis can be successfully applied "to optimize cellulosic fermentation". A similar concern is in fact also raised by reviewer #2, who feels that this aspect is 'essential' albeit leaving it open on whether this should be addressed in this study. We do agree with reviewer #1 that additional evidence would indeed be necessary to support the main claim expressed in the title of this study and would considerably improve the impact of this work.

We would also kindly remind you to deposit your raw MS data in one of the major public repositories and to include the respective accession/hashtag/link in Materials & Methods.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favourable.

Yours sincerely,

Editor
Molecular Systems Biology

REFEREE REPORTS

Reviewer #1 (Remarks to the Author):

General Comments:
The presented manuscript does a thorough job of describing the growth and associated proteome of C. phytofermentans when cultured on different substrates. The title and general overarching presentation are a bit misleading as this manuscript is really a characterization and interpretation of proteomic data, not the development of a novel design strategy. One general comment is that almost all of the Figures should be streamlined and/or truncated. Most subpanels of each Figure are mentioned in a cursory manner and do not substantially contribute to the overall message.

Specific Comments:

1) The main information of this paper can largely be presented using Figure 2, Figure 5A, C-E, Figure 7, and Table 1.

2) Figure 2C - Growth on Cellulose
It is natural that OD cannot be used as a measure of cell culture density when using a solid substrate such as Whatman paper strips for cellulose, but the use of cellulose consumed as a proxy for cell growth is not ideal. The authors should do a more direct measurement such as a Bradford protein assay as the protein content can be correlated more directly to cell growth.

3) Figure 5A
To reduce some potential confusion, the authors may consider changing the bottom-most label to describe the enzymatic activity being tested, not the substrate used for testing.

4) Figure 6
This Figure is largely non-informative and should be removed.

5) Figure 7 - Model
To a large extent the main point of the data and results are presented in this single Figure. The authors present this Figure as the culmination of their analyses and a model of cellulolytic fermentation for C. phytofermentans. The statements that the authors make based off of this model/ set of results are largely self-evident. Figure 7 is basically a super-position of proteomic data information onto a metabolic map. What is the significance of these findings and where is the follow-up evidence that the suppositions proposed by the authors to improve cellulosic ethanol production would actually work?

6) Page 11 lines 312-314
"C. phytofermentans ferments cellulose to ethanol with high specificity (Fig 2F), supporting that accelerating ethanol production rates will give larger gains than streamlining end products."
Either remove or reword this sentence. Of course accelerating ethanol production rates would give larger gains in ethanol production.

7) Figure descriptions in text and Figure legends require elaboration/modification
The Figure 3A Venn diagram is potentially interesting, but the description of these results is terse.

The addition of the secreted protein dashed lines make the diagram much more difficult to read. The authors may want to simplify the Figure and highlight key proteomic differences in the text.
Figure 6 legend - "Nodes are proteins (circles) or yellow diamonds (KEGG/CAZy categories);" Fix this. Nodes are not proteins or yellow diamonds. Nodes are proteins or KEGG/CAZy categories.

Reviewer #2 (Remarks to the Author):

Review of MSB manuscript submission MSB-10-2385 "An Integrated proteomic-based strategy to optimize cellulosic fermentation" by Tolonen et al.

The paper provides a comprehensive strategy to investigate and potentially optimize biological processes for increasing yields of biofuels. The paper is well-written and the authors present convincing evidence to support the major claims (albeit not confirmed independently), therefore publication is recommended after revision addressing the points below.

The main comments that should be addressed in minor revision include:

- How was the ReDi labeling procedure evaluated to ensure accuracy (eg. no experimental evidence is presented for claims on page 3, lines 81-83)? I do not believe mRNA measurements are sufficient to show the strategy performs adequately. Were any other experiments performed to support the author's claims? For example, comparison to SILAC or quantitation of samples with known concentrations of selected proteins?

- Supporting information provides a wealth of important information for the manuscript, however as presented, supporting information is difficult to follow and requires the addition of brief experimental details/discussion to guide the reader through the numerous figures and tables. Also, supplementary figures are not numbered in the sequence in which they are presented in text (eg. Jump from Figure S1 to Figure S4)

- Introduction, authors omit discussion of any other proteomic strategies presented in literature so they do not place the current study in the context of other competing approaches in the field (for example, papers by Zhao and Bai, Journal of Biotechnology, 2009, 144, 23-30 Ito et al, Current Proteomics, 2010, 7, 121-134, Janssen et al, Applied Microbiology and Biotechnology, 2010, 87, 2209 should all be mentioned). These authors are not the first to propose the use of proteomics to guide optimization of biofuel production, and introduction should clearly reflect this.

- Lines 99-110 authors state that ethanol yields for glucose were >95% but in Figure S4, the yield at 160mM glucose is ~50%. Please explain and revise the statement in the main text accordingly.

- lines 131-133, for the comparison of treatments, why weren't the cells in the same growth stage selected for analysis? The experiment, as performed, makes it difficult to compare the treatments in straight-forward manner as the observed differences could be due to growth rate variation as stated by authors. Why was such experimental design selected in the first place? In light of different growth stages, how valid is subsequent interpretation of hemicellulose versus cellulose proteomes presented in lines 253-269?

- Tables S5 and S6? How do the authors explain the presence of the proteins not found in lysate but found in supernatant?

- Table 1 - the meaning of SignalP-NN column is unclear from Table caption. Why was significance threshold set at 0.45?

- Experimental details regarding how ethanol, acetate and glucose were measured are missing from the manuscript/supplementary information and should be included. (lines 347-349, information provided is insufficient to replicate the experiments) Did the authors ensure separation of glucose from other sugars? Example chromatograms should be included in SI.

- Were any experiments undertaken to validate model presented in Figure 7? It is essential, but arguably for a separate follow up study, to see how well the protein candidates for optimization panned out.

- Figure S12 is missing annotations

- Lines 387-389, insufficient information regarding LC-MS provided to replicate the experiments. MS settings? LC conditions? Gradient? Length of gradient? Flow rate? Injection volume?

- Suggest to add replicate for one treatment to Figure 3D. The replicates that were performed - were they only technical replicates or biological replicates?

- Figure S9 correlation coefficients are missing from the caption

- Table S1 - why are replicates not included in this Table?

Reviewer #3 (Remarks to the Author):

Ln 127 add by before APEX.

Ln 211 destratification is not the best term and it is not certain that endocellulases initiate cleavage. In Table 1 need to make clear what is shown in the columns labelled cellulose and hemicellulose and the values should be rounded off to reflect the errors ie there are too many significant figures.

It should be stated that the current rate of cellulose hydrolysis needs to be greatly increased to make this organism useful for biofuel production.

Over all this is an impressive paper.

1st Revision - authors' response 03 November 2010

Point by point response:

While reviewer #2 (expertise: systems biology, proteomics) and reviewer #3 (expertise: cellulose degradation) are rather supportive, reviewer #1 (expertise: systems biology, metabolic engineering) is more reserved. One of the major concerns expressed by reviewer #1 (point 5) refers to the current lack of evidence demonstrating that the systems-level insights (summarized in Fig 7) derived from the present analysis can be successfully applied "to optimize cellulosic fermentation". A similar concern is in fact also raised by reviewer #2, who feels that this aspect is 'essential' albeit leaving it open on whether this should be addressed in this study. We do agree with reviewer #1 that additional evidence would indeed be necessary to support the main claim expressed in the title of this study and would considerably improve the impact of this work.

Yes, we acknowledge that while our systems biology approach identified promising genetic targets to engineer microbes for improved cellulosic bioconversion, we have not yet engineered an optimized strain. Our revised manuscript is entitled "Proteome-wide systems analysis of a cellulosic biofuel producing microbe" to better reflect our systems biology focus. We also highlight in the text that these enzymes are "promising" (line 86) targets to "potentially" (line 288) optimize cellulosic fermentation, but they are not yet definite. We have recently developed tools to over-express and inactivate genes in C. phytofermentans (Tolonen AC et al, 2009). The current focus of our studies is to experimentally manipulate expression of key enzymes identified by this study. These metabolic engineering experiments will the the focus of follow-up studies; here we focus on proteomics-based systems biology.

We would also kindly remind you to deposit your raw MS data in one of the major public repositories and to include the respective accession/hashtag/link in Materials & Methods.

We have uploaded our mass spectrometry .RAW files to the proteomecommons.org repository, which is described on line 407 of the main text. They will be made public upon publication of this manuscript. The hash key to access the data is: 3n+ +Ey/FMTRCCMvwxPAdNuSTL2VVGaWyODpYlSLqyk24mYzCVhaLcFsvcdRvtDi83gNOKYUh PPn84z LjRdBjWA2kUIgAAAAAAABFUw==

Reviewer #1 (Remarks to the Author):
General Comments: The presented manuscript does a thorough job of describing the growth and associated proteome of C. phytofermentans when cultured on different substrates. The title and general overarching presentation are a bit misleading as this manuscript is really a characterization and interpretation of proteomic data, not the development of a novel design strategy. One general comment is that almost all of the Figures should be streamlined and/or truncated. Most subpanels of each Figure are mentioned in a cursory manner and do not substantially contribute to the overall message.

Response: We appreciate your comments regarding the thoroughness of our study. We agree that the title and presentation of our original manuscript needed to be changed to better reflect our analyses and findings. In our revised manuscript, the text and title have been changed the title to emphasize the importance of our proteomic-based approach to understand the systems biology of this cellulosic biofuel producing microbe. The title now reads "Proteome-wide systems analysis of a cellulosic biofuel producing microbe". We respond to your suggestions regarding the figures below.

Specific Comments:
1) The main information of this paper can largely be presented using Figure 2, Figure 5A, C-E, Figure 7, and Table 1.

Response: We agree that the figures are information-rich and some may not be needed from the perspective of an organism design strategy. However, since the focus of our revised manuscript is to present a comprehensive systems analysis of a cellulosic biofuel producing microbe, we feel that all figures contain important information. Fig 1 gives an overview of how the different data types fit together to form the strategy used in this paper to study cellulosic bioconversion, which we think puts our methods in perspective for a general audience. Fig 3 is needed to evaluate the quality and completeness of the ReDi proteomic data, which is critical to our proteomics-based systems biology and was a focus of reviewer 2. Analysis of the secretome and secretory mechanisms in Fig 4 are of particular interest because the extracellular cellulolytic enzymes are secreted by these pathways. This data is applied to propose efficient secretory signals in C. phytofermentans (Fig 4 B-C) to facilitate secretion of heterologous proteins and improve secretion of endogenous proteins. To our knowledge, this is also the first comprehensive validation of the SignalP3.0 neural network (Fig 4A), which is widely used to predict secreted proteins. Fig 6 is gives a proteome-wide view of the expression changes on cellulose, revealing novel changes related to biomass utilization such as up-regulation of tryptophan and nicotinamide biosynthesis and repression of fatty acid synthesis and motility proteins. These findings highlight the value of a systems level approach and are included in the abstract as some of the most important findings of the study.

2) Figure 2C - Growth on Cellulose It is natural that OD cannot be used as a measure of cell culture density when using a solid substrate
such as Whatman paper strips for cellulose, but the use of cellulose consumed as a proxy for cell growth is not ideal. The authors should do a more direct measurement such as a Bradford protein assay as the protein content can be correlated more directly to cell growth.

Response: Cellulose consumption (Fig 2C) and ethanol formation (Fig 2F) were shown to highlight

the efficient direct conversion of cellulose to ethanol. We agree that cellulose consumption is not a good measure of cell growth. For this reason, we directly quantified cell growth as colony forming units on plates (lines 101-102).

3) Figure 5A To reduce some potential confusion, the authors may consider changing the bottom-most label to describe the enzymatic activity being tested, not the substrate used for testing.
Response: The bottom-most labels of Fig 5A have been changed to describe the enzyme activities being tested. Specifically, 'cellulose' has been changed to 'cellulase' and 'hemicellulose' has been changed to 'hemicellulase'.

4) Figure 6 This Figure is largely non-informative and should be removed.

Response: Please see our response to comment 2 describing the importance of this figure to show specific, novel findings revealed by the proteome-wide expression changes on cellulosic substrates.

5) Figure 7 - Model To a large extent the main point of the data and results are presented in this single Figure. The authors present this Figure as the culmination of their analyses and a model of cellulolytic fermentation for C. phytofermentans. The statements that the authors make based off of this model/ set of results are largely self-evident. Figure 7 is basically a super-position of proteomic data information onto a metabolic map. What is the significance of these findings and where is the follow- up evidence that the suppositions proposed by the authors to improve cellulosic ethanol production would actually work?

Response: We regret that the reviewer feels that Fig 7 is self-evident, but we respectfully disagree. Our mass spectrometry data did not come interpreted on a metabolic map and determining how these data fit into a metabolic network required considerable effort on our part. Furthermore, this network includes specific adaptations used by C. phytofermentans for efficient anaerobic fermentation that were not previously known. For example, our proteomic data revealed the high expression of alternative glycolytic enzymes that are reversible and use pyrophosphate (PPi) as a phosphate donor. As PPi-dependent glycolytic enzymes have been shown to increase ATP yields in anaerobic eukaryotes (Slamovits and Keeling, 2006), we propose that this is also an important adaption to improve ATP yields in C. phytofermentans. This is but one example and we hope it clarifies our belief in the importance of this Figure.
We strongly agree with the reviewer that follow-up work is critical to demonstrate that the findings from this study can be applied to improve the organismís biofuel producing capability. This work is, in fact, in progress. We have recently developed methods for experimental manipulation of gene expression in C. phytofermentans (Tolonen AC et al, 2009). As the reviewer suggests, our current focus is to metabolically engineer C. phytofermentans to examine which of the genetic targets revealed by this study can be used to improve cellulosic fermentation. These will be the basis of follow-on articles, but here we have preferred to focus on systems biology.

6) Page 11 lines 312-314 "C. phytofermentans ferments cellulose to ethanol with high specificity (Fig 2F), supporting that accelerating ethanol production rates will give larger gains than streamlining end products." Either remove or reword this sentence. Of course accelerating ethanol production rates would give larger gains in ethanol production.

Response: This sentence has been changed to "C. phytofermentans ferments cellulose to ethanol with high specificity (Fig 2F), supporting that accelerating biomass deconstruction will give larger gains than streamlining end products."

7) Figure descriptions in text and Figure legends require elaboration/modification The Figure 3A Venn diagram is potentially interesting, but the description of these results is terse. The addition of the secreted protein dashed lines make the diagram much more difficult to read. The authors may want to simplify the Figure and highlight key proteomic differences in the text.

Response: Removing the supernatant samples would simplify the Venn diagram (Fig 3A), but it would also reduce its information content and undercut the completeness of the dataset. Solid ellipses corresponding to the culture lysates show the total number and the overlap in protein identification among the carbon sources. By adding the dashed ellipses for the supernatant proteins, we also show the fraction of the total proteome that is exported from the cell (the secretome) and how the secretome changes as a function of carbon source.
Figure 6 legend - "Nodes are proteins (circles) or yellow diamonds (KEGG/CAZy categories);" Fix this. Nodes are not proteins or yellow diamonds. Nodes are proteins or KEGG/CAZy categories.

Response: Thank you for catching this typo. This statement has been changed to "Nodes are proteins (circles) or KEGG/CAZy categories (yellow diamonds)" on lines 710-712.

Reviewer #2 (Remarks to the Author):

Review of MSB manuscript submission MSB-10-2385 "An Integrated proteomic-based strategy to optimize cellulosic fermentation" by Tolonen et al.
The paper provides a comprehensive strategy to investigate and potentially optimize biological processes for increasing yields of biofuels. The paper is well-written and the authors present convincing evidence to support the major claims (albeit not confirmed independently), therefore publication is recommended after revision addressing the points below.

Response: We thank the reviewer for praising our study as comprehensive with convincing evidence to support our claims. We respond to the reviewerís specific suggestions below.

The main comments that should be addressed in minor revision include:
- How was the ReDi labeling procedure evaluated to ensure accuracy (eg. No experimental evidence is presented for claims on page 3, lines 81-83)? I do not believe mRNA measurements are sufficient to show the strategy performs adequately. Were any other experiments performed to support the author's claims? For example, comparison to SILAC or quantitation of samples with known concentrations of selected proteins?

Response: In response to the reviewer's comment, we have included new data to evaluate the accuracy, precision, and reproducibility of ReDi labeling. The methods are detailed in the ReDi Protein Quantification section of the Supplementary Information and the results are shown in Fig S11 (see below) and Table SII. Briefly, differentially labeled (heavy (H) or light (L)) protein samples were mixed at different ratios and analyzed by mass spectrometry. Fig S11 compares replicate samples to show that ReDi quantifications reproducibly and accurately reflect the ratios at which the H and L samples were mixed.

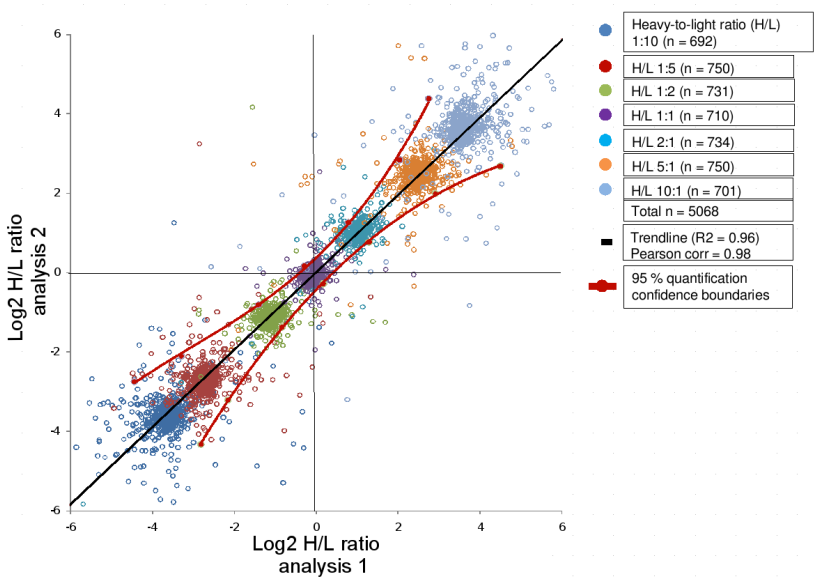

Fig S11 ReDi labeling evaluated by comparing protein abundances (log2 H/L ratios) of replicate experiments for H and L samples mixed at the following ratios: 1H:10L, 1H:5L, 1H:2L, 1H:1L, 2H:1L, 5H:1L, and 10H:1L. The R2 value of a linear trend line through all data points was 0.96, the Pearson correlation was 0.98. We also determined reproducibility confidence boundaries defined as the log2 H/L difference between the H/L ratio measured for any protein in the duplicate measurements. We calculated these boundaries by measuring the absolute perpendicular distance from the trend line within which 95 % of the data points were found for for each sample. These boundaries are shown as a red lines.

- Supporting information provides a wealth of important information for the manuscript, however as presented, supporting information is difficult to follow and requires the addition of brief experimental details/discussion to guide the reader through the numerous figures and tables. Also, supplementary figures are not numbered in the sequence in which they are presented in text (eg. Jump from Figure S1 to Figure S4)

Response: We have expanded the Supp Info to include additional experimental details/discussion as suggested by the reviewer below. We have also modified the main text to cite Supplementary figures 2 and 3 (line 99). All supplementary figures and tables are now referenced in the main text.

Introduction, authors omit discussion of any other proteomic strategies presented in literature so they do not place the current study in the context of other competing approaches in the field (for example, papers by Zhao and Bai, Journal of Biotechnology, 2009, 144, 23-30 Ito et al, Current Proteomics, 2010, 7, 121-134, Janssen et al, Applied Microbiology and Biotechnology, 2010, 87, 2209 should all be mentioned). These authors are not the first to propose the use of proteomics to guide optimization of biofuel production, and introduction should clearly reflect this.

Response: We thank the reviewer for pointing out these recent reviews by Zhao and Bai 2009 and Ito et al, 2010, which we now cite. We describe on lines 309-311 how our proteome-wide study builds upon previous proteomic work on the cellulosomes in other cellulolytic clostridia (Gold and Martin, 2007; Raman et al, 2009; Blouzard et al, 2010). Janssen et al, 2010 is an excellent study on C. acetobutylicum, but we did not include this reference because this organism does not metabolize cellulosic substrates.

Lines 99-110 authors state that ethanol yields for glucose were >95% but in Figure S4, the yield at 160mM glucose is ~50%. Please explain and revise the statement in the main text accordingly.

Response: We have revised our description of Supplementary Fig S4, now Fig S6. This figure shows that ethanol yields were >95% at in media with initial glucose concentrations of 16, 32, and 80 mM

glucose. Lower ethanol yields in media with an initial glucose concentration of 160 mM were included to show the limitation of using GS2 medium at higher glucose concentrations. Cells become limited by another nutrient, resulting in lower ethanol yields. High ethanol yields with initial glucose concentrations >80 mM require different media or additional yeast extract in GS2 medium. We have expanded our explanation of the ethanol yields at 160 mM glucose in the legend of Fig S6.

Lines 131-133, for the comparison of treatments, why weren't the cells in the same growth stage selected for analysis? The experiment, as performed, makes it difficult to compare the treatments in straight-forward manner as the observed differences could be due to growth rate variation as stated by authors. Why was such experimental design selected in the first place? In light of different growth stages, how valid is subsequent interpretation of hemicellulose versus cellulose proteomes presented in lines 253-269?

Response: These experiments were designed to quantify global changes in the proteome during growth on hemicellulose and cellulose relative to glucose. Cells were in the same growth stage in all treatments in that they were actively growing in nutrient replete media and metabolizing the substrates to ethanol (i.e. log phase growth) when samples were taken for proteomics (lines 92-94). Fig 2 shows how metabolizing these different carbon sources entailed physiological changes such as growth rate and loss of cell motility. Thus, differences in the proteome reflect both proteins directly related to substrate utilization (i.e. cellulolytic enzymes) and those indirectly linked to cellulose metabolism (i.e. down-regulation of transcription and translation). We have revised the text (lines 130-133) to clarify this matter.

Tables S5 and S6? How do the authors explain the presence of the proteins not found in lysate but found in supernatant?

Response: Both supernatant treatments include proteins that were observed at low abundance in the supernatant, but not in the culture lysate. Because of the lower number of proteins in the supernatant samples, we were able to get deeper coverage and observe proteins that were below detection in the more complex culture lysate samples. We have added a paragraph to the "Filtering Intracellular Proteins" description in the Secretome section of the Supplementary Information to clarify this matter.

- Table 1 - the meaning of SignalP-NN column is unclear from Table caption. Why was significance threshold set at 0.45?

Response: The 0.45 threshold we used for D-value cutoffs was computed and reported by the signalP3.0 neural network (NN) software itself, and is not a threshold we developed on our own. Based on our understanding of the SignalP references (see http://www.cbs.dtu.dk/services/SignalP/), the performance of a variety of signal peptide classifiers was evaluated by five-fold cross-validation to compute a "discrimination" score for each classifier. This score is described as a correlation coefficient between the classifier output and (we presume) the indicator variables corresponding to the left-out training sets. D-values are identified as the best-performing classifier scores, and in the case of Table I (Bendtsen JD, Nielsen H, von Heijne G, and Brunak S. (2004). Improved prediction of signal peptides: J. Mol. Biol., 340:783-795), D-values computed for gram positive microbes achieve a correlation coefficient of 0.98. Beyond this kind of information, the authors do not report probabilistic significances such as sensitivities or specificities, and we see no way of computing one from the correlation coefficients beyond noting that R^2 (in this case .96) might be construable as the fraction of explained variance. We suspect the D-value of 0.45 likely comes from training the NN on indicator values of 0 and 1 for the training set, followed by adjustment of an initial 0.5 threshold for NN output scores to minimize classifier error (a standard NN technique). However, this level of detail is not provided by the authors and represents only our best interpretation.

Experimental details regarding how ethanol, acetate and glucose were measured are missing from the manuscript/supplementary information and should be included. (lines 347-349, information

provided is insufficient to replicate the experiments) Did the authors ensure separation of glucose from other sugars? Example chromatograms should be included in SI.

Response: The supplementary information has been expanded to include example HPLC chromatograms of glucose consumption (Fig S3) and accumulation of fermentation products (Fig S4) (see below). We also added a thorough description of how compounds were quantified by HPLC using standard curves generated with solutions of known concentration. We have re-worded the Fig 2 legend to clarify that glucose consumption was quantified only in the glucose treatment. Sugars in the hemicellulose and cellulose treatments were not measured because they did not produce distinct, detectable HPLC peaks.

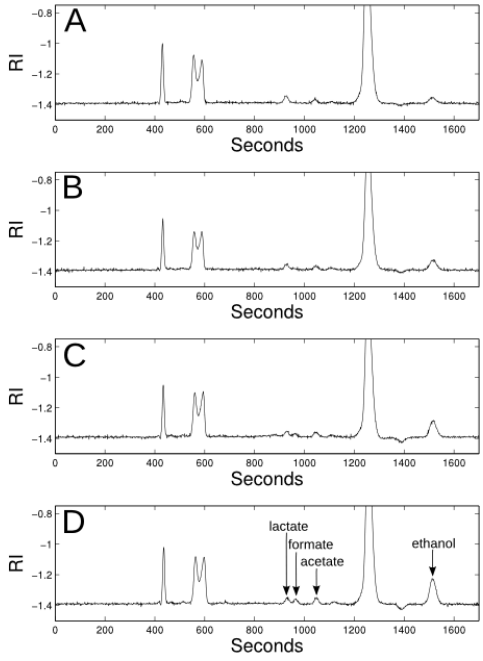

Fig S4 HPLC chromatograms of one of the cellulose culture in the time series shown in Fig 2F of the main text. Measurements were taken on A 7 days, B 14 days, C 21 days, and D 28 days. Chromatograms show accumulation of fermentation products: ethanol (1500 s), acetate (1050 s), formation (960 s), and lactate (880 s). Ethanol and acetate were quantified and shown in Fig 2F.

- Were any experiments undertaken to validate model presented in Figure 7? It is essential, but arguably for a separate follow up study, to see how well the protein candidates for optimization panned out.

Response: We agree that investigating how the metabolic model presented in figure 7 can be applied to improve biofuel production is of high importance, and this is the current focus of our work. We have recently developed methods for over-expression and gene inactivation in C. phytofermentans (Tolonen AC et al, 2009). We are applying these tools to examine which of the genetic targets revealed by this study can be used to improve cellulosic fermentation. These will be the basis of follow-up articles, but here we have focused on proteomics-based systems biology.

Figure S12 is missing annotations

Response: We have added pathway annotations to this figure, which is now Fig S14.
- Lines 387-389, insufficient information regarding LC-MS provided to replicate the experiments. MS settings? LC conditions? Gradient? Length of gradient? Flow rate? Injection volume?
Response: We have expanded the Quantitative Proteomics methods section to describe this

information as part of a detailed description of the LC-MS/MS conditions (lines 393-407).

Suggest to add replicate for one treatment to Figure 3D. The replicates that were performed - were they only technical replicates or biological replicates?

Response: Fig 3D compares proteome-wide ReDi quantifications of a glucose culture to another glucose culture (biological replicate), a hemicellulose culture, and a cellulose culture. Thus, Fig 3D shows a replicate for one treatment (glucose), which is discussed in the main text as how "we observed 94% of proteins expressed within 2-fold levels for differentially-labeled duplicate glucose cultures" (lines 145-147). Fig 3E is a scatter plot giving a more detailed view of ReDi quantifications for biological replicates of cellulose vs glucose cultures showing "global expression changes on cellulose versus glucose were well correlated (r2=0.82) in replicate pairs of cultures" (lines 150-151).

Figure S9 correlation coefficients are missing from the caption

Response: Fig S9 of the original manuscript, which showed high correlation of APEX between peptides identified by the Target-Decoy method and the Trans-Proteomic Pipeline, has been omitted because this methods comparison is no longer a focus of the manuscript.

Table S1 - why are replicates not included in this Table?

Response: Supplementary Table SI shows the APEX values from the samples that were used in the analyses throughout this study. The correlation between these APEX and replicate cultures were (r2=0.94-0.97) for all three treatments (Supplementary Fig S10).

Reviewer #3 (Remarks to the Author):

Ln 127 add by before APEX.

Response: Thank you for catching this typo. We have added "by" before "APEX" on line 127.

Ln 211 destratification is not the best term and it is not certain that endocellulases initiate cleavage.

Response: We have replaced " Initial destratification" with "Depolymerization" (line 217).

In Table 1 need to make clear what is shown in the columns labelled cellulose and hemicellulose and the values should be rounded off to reflect the errors ie there are too many significant figures.

Response: We have reduced the APEX values in Table 1 to a single significant figure.

It should be stated that the current rate of cellulose hydrolysis needs to be greatly increased to make this organism useful for biofuel production.

Response: We have revised the final paragraph of the Introduction (lines 71-87) to further emphasize how cellulosic biofuels by microbes such as C. phytofermentans will require improved cellulolysis.

Over all this is an impressive paper.
Response: Thank you.

Acceptance letter 30 November 2010

Thank you again for sending us your revised manuscript. We have now heard back from the reviewers who are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

IMPORTANT:
Before we can transfer your manuscript to production, we would kindly ask you to include the information requested by reviewer #2 in supplementary information and send us the updated zip file so that we can upload it on our website.

Thank you very much for submitting your work to Molecular Systems Biology.

Sincerely,
Editor
Molecular Systems Biology

Reviewer #1 (Remarks to the Author):

The authors have adequately addressed the majority of the initial concerns with the manuscript. The revised manuscript is improved in clarity, focus, and presentation from the initial manuscript.

Reviewer #2 (Remarks to the Author):

Revision looks good. They addressed all my comments. The only thing that is still missing is mobile phase conditions for glucose and fermentation products chromatograms (Figures S3 and S4). Everything else looks fine to me.

