ID: ER65257-1038489-0017559
Principal Investigator: Eleftherios Papoutsakis 302-831-8376
Co-PIs: Blake C. Meyers, Kelvin H. Lee, Maciek R. Antoniewicz, Costas D. Maranas, Hongzhan Huang and Cathy H. Wu.
Institution: University of Delaware
Title: Experimental Systems-Biology Approaches for Clostridia-Based Bioenergy Production
SC Division: SC-23.2
Program Manager: Dean A. Cole; Roland Hirsch (BER; Ph: 301-903-9009; Roland.Hirsch@science.doe.gov)

Final project report for the funding period of 9/1/12 to 2/28/2015 (three years with a 6-month no-cost extension)

OVERVIEW & PROJECT GOALS

The bottleneck of achieving higher rates & titers of toxic metabolites (such as solvents and carboxylic acids that can used as biofuels or biofuel precursors) can be overcome by engineering the stress response system. Thus, understanding & modeling the response of cells to toxic metabolites is a problem of great fundamental and practical significance. In this project, our goal is to dissect at the molecular systems level and build models (conceptual and quantitative) for the stress response of *C. acetobutylicum* (*Cac*) to its two toxic metabolites: butanol (BuOH) & butyrate (BA). Transcriptional (RNAseq and microarray based), proteomic & fluxomic data and their analysis are key requirements for this goal.

Transcriptional data from mid-exponential cultures of *Cac* under 4 different levels of BuOH & BA stress was obtained using both microarrays (Papoutsakis group) & deep sequencing (RNAseq; Meyers & Papoutsakis groups). These two sets of data do not only serve to validate each other, but are also used for identification of stress-induced changes in transcript levels, small regulatory RNAs, & in transcriptional start sites. Quantitative proteomic data (Lee group), collected using the iTRAQ technology, are essential for understanding of protein levels and turnover under stress & the various protein-protein interactions that orchestrate the stress response. Metabolic flux changes (Antoniewicz group) of core pathways, which provide important information on the re-allocation of energy & carbon resources under metabolite stress, were examined using ^{13}C-labelled chemicals. Omics data are integrated at different levels and scales. At the metabolic-pathway level, omics data are integrated into a 2nd generation genome-scale model (GSM) (Maranas group). Omics data are also integrated using bioinformatics (Wu & Huang group), whereby regulatory details of gene & protein expression, protein-protein interactions & metabolic flux regulation are incorporated. The PI (Papoutsakis) facilitated project integration through monthly meeting & reports, conference calls, and collaborative manuscript preparation. The five groups collaborated extensively and made a large number of presentations in national and international meetings. It has also published several papers, with several more in

the preparation stage. Several PhD, MS and postdoctoral students were trained as part of this collaborative and interdisciplinary project.

SUMMARY OF RESULTS & FINDINGS

The Papoutsakis group

Summary

The Papoutsakis group was responsible for the overall project coordination, resource and funds management, and reporting. Most importantly, it had the necessary expertise regarding the experimental design of the microbial stress experiments and the collection of biological samples (RNA and proteins) for the omics data collection. It also carried out most of the omics analysis at the RNA level (RNAseq and microarray analyses; for RNAseq analysis, the Meyers group played a critical role method development and training), as well as the development of the omics analysis pipeline. The group carried out an exceptionally large number of well-controlled culture experiments with the two metabolite stressors (butanol and butyrate), collected all the RNA omics data, coordinated proteome analysis with the Lee group, bioinformatics analyses with the Wu and Huang groups, 13C flux analysis and model building with the Antoniewicz group, and the model building efforts with the Maranas group. It also coordinated integrated omics (RNA and proteomic) analyses.

Methods and Results

We carried out cultures of *Cac* in 4 L bioreactors for the 4 stress levels for each metabolite to collect samples (at several time points: 0, 15, 30, 45, 60 & 75 min post stress) for for transcriptomic (microarray and RNAseq) & proteomic analyses. Each set of experiments was carried out with 3-6 biological replicates, and RNAseq analysis was carried out at both the standard method as well as with strand-specific RNAseq. These data were used to build several stories described below as organized in publications (published, in press or in preparation).

The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum (in collaboration with the Meyers group. Publication: *BMC Genomics* 2013, 14:849). Small non-coding RNAs (sRNA) are emerging as major components of the cell's regulatory network, several possessing their own regulons. A few sRNAs have been reported as being involved in general or toxic-metabolite stress, mostly in Gram- prokaryotes, but hardly any in Gram+ prokaryotes. Significantly, the role of sRNAs in the stress response remains poorly understood at the genome-scale level. Using RNA deep sequencing (RNA-seq) we examined the sRNome of C. acetobutylicum in response to the native but toxic metabolites, butanol and butyrate. 50% of the RNA-seq reads mapped to genome outside annotated ORFs, thus demonstrating the richness and importance of the small RNome. We used comparative expression analysis of 113 sRNAs we had previously computationally predicted, and of annotated mRNAs to set metrics for reliably identifying sRNAs from RNA-seq data, thus discovering 46 additional sRNAs. Under metabolite stress, these 159 sRNAs displayed distinct expression patterns (**Figure 1**), a select number of which was verified by Northern analysis. We identified stress-related expression of sRNAs affecting transcriptional (6S, S-box & solB) and translational (tmRNA & SRP-RNA) processes, and 65 likely targets of the RNA chaperone Hfq.

Transcription factors and genetic circuits orchestrating the response of Clostridium acetobutylicum to butanol and butyrate stress (in collaboration with the Wu and Huang group. Publication: *BMC Systems Biology* 2013, 7:120). Chemical and in particular toxic metabolite stress engages the general stress response as well as specialized programs. However, the transcriptional regulatory network underlying the stress response remains largely unexplored at the systems level. We generated a total of 192 individual set of high-resolution microarray data examining the transcriptional changes in *C. acetobutylicum* in response to three levels of chemical stress from the native metabolites, butanol and butyrate. We identified 164 significantly differentially expressed transcription regulators and detailed the cellular programs that are linked to either general stress response or stressor-specific response. Comparisons to several previous studies were made and allowed new insight derived, e.g., stressor dose-dependent and culture condition-dependent genes and pathways are revealed. Mining of our new comprehensive data and conduction of comparative genomic analyses allowed the construction of a detailed picture of the genetic circuitry underlying the stress response. In particular, the regulons involving stress-related transcription factors such as HrcA, CtsR, LexA, Rex and PerR are defined, together with those for select amino acid and purine metabolism (i.e., ArgR, HisR, CymR and PurR) that were found to be stress responsive.

Capturing the response of Clostridium acetobutylicum to chemical stressors using a regulated Genome-scale metabolic model (in collaboration with the Maranas group. Publication: *Biotechnology for Biofuels*, 7(1), 144). This paper and story describe the construction, validation, and integration of a GSM model for *C. acetobutylicum* ATCC 824, *i*Cac802, with experimental gene expression data to predict the cellular responses to two chemical stressors: butanol and butyrate. More details are provided under the report of the Maranas group.

Integrative proteomic and transcriptomic analysis of Clostridium acetobutylicum's response to butanol and butyrate stress reveals complex post transcriptional regulation (in collaboration with the Lee group. Publication: *Biotechnology for Biofuels* 2015*, in press*, DOI: 10.1186/s13068-015-0260-9). Integrative analysis of proteomic and RNAseq data may provide novel insights into post-transcriptional regulation of *Clostridium acetobutylicum* under both stress and health. The identified iTRAQ-based quantitative stress proteome is made up of 616 proteins with a15% genome coverage (**Figure 2**). The differentially expressed proteome correlated poorly with the corresponding differential RNAseq transcriptome. Up to 31% of the differentially expressed proteins under stress displayed patterns opposite to those of the transcriptome (see, e.g., **Figure 3**), thus suggesting significant post-transcriptional regulation. The differential proteome of the translation machinery suggests that cells employ a different subset of ribosomal proteins under stress. Several highly upregulated proteins but with low mRNA levels possessed mRNAs with long 5'UTRs and strong RBS scores, thus supporting the argument that regulatory elements on the long 5'UTRs control their translation.

Strand-specific RNAseq analysis of Clostridium acetobutylicum enables accurate transcriptome assembly revealing new genes and the transcriptional intricacies of large and small RNAs, in health and under stress (Manuscript in preparation).
Next Generation Sequencing (NGS) of RNA can produce hundreds of millions of sequenced reads. RNAseq relies on the proportionality between expression levels and the number of sequenced reads. Transcriptome assembly is rarely used in bacterial RNA-seq studies; frequently, depth alone informs the researchers decision on transcript boundaries with no

discussion of background signals. We carried out a transcriptome assembly from *Clostridium acetobutylicum* using high-depths strand-specific RNAseq data. The resulting assembly was optimized with functional transcriptomic metrics and previous work in *C. acetobutylicum* in mind. Cross validation of this assembly demonstrates the robustness of the integrative assembly and curation method used here. We identified several new transcripts as well TSSs, 5'UTRs, 3'UTRs and several other novel molecular features of the assembled transcriptome.

DNA methylation in Clostridium acetobutylicum is altered under metabolite stress, in stationary phase and in the absence of the megaplasmid (in collaboration with the Wu and Huang group. Publication: in preparation). With the development of the PacBio SMRT (Single Molecule Real Time) sequencing technology, it is now feasible to study the whole genome epigenetic changes in the bacterial genome. In this study, we identified, for the first time, the methylation pattern in *C. acetobutylicum* in response to two metabolite stressors (butyrate and butanol), between two strains (WT and M5; M5 lacks the pSOL1 megaplasmid that contains the solventogenic genes), and between two growth phases (exponential and stationary phase) using PacBio's RS II platform.

Conclusions:

1. Our results support an important role for sRNAs for understanding the complexity of the regulatory network that underlies the stress response in Clostridium organisms, whether related to normophysiology, pathogenesis or biotechnological applications.

2. Using a large set of temporal transcriptional data, we were able to successfully build a regulatory network model for the general and specialized metabolite stress response in *C. acetobutylicum*. A large part of this network is applicable to other *Clostridium* organisms. This network will facilitate the construction of genome-scale models with added regulatory-network dimensions to guide future development of tolerant and productive strains.

3. The regulation placed on the model for the two stresses using CoreReg identified differences in the respective responses, including distinct core sets and the restriction of biomass production similar to experimental observations. Given transcriptomic data the CoreReg method can be used to predict an organism's response to other stressors by identifying core sets of reactions whose down-regulation propagates through stoichiometry to the remaining metabolic network causing flux changes consistent with experimentally observed trends.

4. The integrative proteomic-transcriptomic analysis demonstrated complex expression patterns of a large fraction of the proteome. Such patterns could not have been detected with one or the other omics analyses. Our analysis proposes the involvement of specific molecular mechanisms of post-transcriptional regulation to explain the observed complex stress response.

5. We completed a transcriptome assembly from *Clostridium acetobutylicum* using high-depths strand-specific RNAseq data. As a result, we identified several new transcripts as well TSSs, 5'UTRs, 3'UTRs and several other novel molecular features of the assembled transcriptome.

6. We found that a set of methylation patterns is always present with almost 100% methylation under all culture conditions. In contrast, two other methylation motifs are under-methylated and

display higher methylation under metabolite stress, but lower methylation in stationary phase or in strain M5. More details are provided under the Wu and Huang report below.

Meyers Group

Summary: The Meyers group focused on transcriptional analysis of *Clostridium acetobutylicum*, generating RNA data for use within the project. Using RNA deep sequencing (RNA-seq), we demonstrated a rich abundance and diversity of transcripts and small RNAs in *C. acetobutylicum*, regulated in response to the metabolites butanol and butyrate. We found 159 sRNAs that displayed distinct expression patterns, validated them, and identified stress-related expression affecting transcriptional and translational processes.

Goals: sRNAs in bacteria span a wider size range between 50 to 500 nts. Regulation of gene expression at post-transcriptional level by sRNAs has been established, so we sought to characterize these on a genome-wide scale in *C. acetobutylicum*.

Methods*:* Using RNA-seq, we aimed to identify sRNAs (previously predicted and novel) that are differentially expressed under butanol and butyrate stress. To do so, we aimed to collect a large set of temporal data, which, based on our experience are more likely to lead to robust discovery outcomes. Cultures of *C. acetobutylicum* were grown in batch mode in 4-L bioreactors up to the mid-exponential phase of growth (O.D ~ 1.0), at which point the cultures were stressed with three different concentrations of butanol and butyric acid, respectively, in three biological-replicate experiments. This work was done with the Papoutsakis lab. Following RNA isolation, mRNA and sRNA enrichment, cDNA generation, adapter ligations and indexing, libraries were deep sequenced using Illumina's second generation HiSeq 2000 with a read length of 50 bp. Sequencing data were obtained for 84 sequenced libraries from samples representing 7 distinct culture conditions with 4 time points and 3 biological replicates each. RNAseq data were analyzed computationally to identify differentially expressed genomic elements (genes, non-coding small RNAs & interoperonic regions) using the DESeq program of the R Bioconductor package. These data are being integrated with the microarray data from the Papoutsakis group for validation & confirmation of differentially regulated loci.

Results*:* We examined the expression profiles of the 159 (113 previously identified and the newly identified 46) sRNAs aiming to identify which are expressed and differentially expressed under the various metabolite-stress conditions. Butyrate stress gave rise to more (45) differentially upregulated sRNAs than butanol stress (33), while butanol stress had more differentially downregulated sRNAs (51) compared to butyrate (44). 42 sRNAs were differentially expressed under both metabolite stresses: 21 were upregulated and 21 were downregulated under both stresses. Although the two metabolite stresses result in differential expression of specific sets of sRNAs that are stress and dose dependent, we found a considerable conservation of expression patterns for the two stressors among these sRNAs, thus suggesting a possible role of these sRNAs in the general stress response.

Expression patterns under metabolite stress of the 159 sRNAs were compared against the non-stressed control cultures (pair-wise & point-by-point) and analyzed using hierarchical clustering. Both butyrate and butanol stress data displayed distinct clusters. The clustered data were analyzed to identify shared regulatory elements, such as promoter sequences and transcription factor binding sites (TFBS) upstream of the sRNAs in the same cluster.

Identification of regulatory elements in the differentially expressed sRNA clusters revealed the presence of both general stress responsive elements (σB) and the more specific oxidative stress response regulators (FNR, ArgR and Rex) supports the clustering of co-regulated stress responsive sRNAs.

Conclusions*:* Our results support an important role for sRNAs for understanding the complexity of the regulatory network that underlies the stress response in *Clostridium* organisms, whether related to normophysiology, pathogenesis or biotechnological applications.

The Lee group

Summary: The Lee developed a workflow for quantitative proteomic analysis of *Clostridia acetobutylicum* (*Cac*) using iTRAQ tags. The workflow was applied to *Cac* cells under either butanol (BuOH) or butyrate (BA) stress, and no stress. A total of 440 and 589 proteins were identified under BuOH and BA stress, respectively. Among these, 149 and 167 proteins had statistically significant differential expressions.

Goals: In this study, iTRAQ label and LC/LC-MS/MS were employed to identify and quantify proteomic changes in *Cac* under BuOH and BA stress.

Methods: *Cac* cells were grown and exposed to no, low, medium and high levels of either BuOH (0, 30, 60 & 90 mM) or BA (0, 30, 40 & 50 mM). *SolRH,* a *Cac* strain with superior tolerance to metabolite stress, were grown at high levels of BuOH (90 mM) and BA (50 mM) stress. Cells were collected at 15, 45 and 75 min after stress exposure. For proteomic analysis, proteins from different stress levels and time points were extracted after lysing cells with sonication with the aid of calcium carbonate; then proteins were digested with trypsin. Digested samples from the same stress levels at the three different time points and a common reference sample were labeled with 4 different iTRAQ tags, then pooled together and subjected to multidimensional liquid chromatography and mass spectrometry (LC/LC-MS/MS). Protein identification and quantification were performed with Protein Pilot v3. Protein expression levels in different levels of metabolite stress were compared to corresponding no stress control using SAM analysis with MeV v4.8 to identify proteins with significantly different expression. Protein identifications were based on 95% confidence or above. Protein expression changes of ≥2.0 fold were considered as significant.

Results: We developed and published a workflow for quantitative proteomic analysis using iTRAQ tags [1]. With this workflow, 440 and 588 unique proteins were identified and quantified under BuOH and BA stress, respectively; 413 proteins were identified under both BuOH and BA stress. Among these, 48, 76 and 75 proteins had significantly differential expressions under low, medium and high BuOH stress, respectively; 55, 64 and 58 proteins had significantly differential expressions under low, medium and high BA stress. In addition, there were proteins detected only under stress conditions, including heat shock proteins (HSPs), UV-stress response proteins, transcriptional regulators, response regulators involved in signal transduction, and chemotaxis proteins. An example of the data collected from BA stressed cells is shown in **Figure 4**, which depicts a COG breakdown of proteins observed in a typical experiment. Proteins were identified

and quantified from nearly every COG category. **Figure 5** shows the COG classification of identified proteins compared to proteins with significantly differential expressions upon BA stress. There are more COG groups with increased expression levels than decreased expression levels. The main COG groups that showed increased expression included COG groups C (energy production and conversion), E (amino acid transport and metabolism), F (nucleotide transport and metabolism), G (carbohydrate transport and metabolism), J (translation, ribosomal structure and biogenesis), O (posttranslational modification, protein turnover, chaperones). Then we worked with the Papoutsakis group and compared the stress proteomic data with the corresponding transcriptomic data from microarray analysis [2] and RNAseq [3] to identify the post-transcriptional regulations. The paper was recently accepted for publication in *Biotechnology for Biofuels* [4].

In the *SolRH* strain, proteomic analysis identified 38 and 41 proteins with significant changes under high levels of BA and BuOH stress, respectively. Under BA stress, the expression levels of 21 and 17 proteins were significantly increased and decreased, respectively. Under BuOH stress, the expression levels of 22 and 19 proteins were significantly increased and decreased, respectively. In contrast to WT, the expression levels of chaperone protein groEL (CA_C2703) and molecular chaperone hsp 18 (CA_C3714) in *SolRH* were significantly decreased under BuOH and BA stress.

To understand the different trends for the chaperone proteins, protein expression levels in *SolRH* were compared to the WT. Under no stress, the expression levels of 58 and 52 proteins in *SolRH* were significantly increased and decreased, respectively; proteins with significant changes constituted 50% of all commonly quantified proteins. Despite the differences between the two strains, protein expression levels under BuOH and BA stress are very similar between WT and *SolRH*, and 95% of commonly quantified proteins did not show significant changes.

Conclusions: A workflow for quantitative proteomic analysis using iTRAQ tags was developed and applied to *Cac*. This method will be applicable to *Cac* studies by any laboratory and is likely useful also to teams studying any Gram-positive organism. With this workflow, we identified and quantified the proteomic changes in *Cac* under BuOH and BA stress.

The Antoniewicz group

***Summary*:** The Antoniewicz group studied the effects of butanol and butyric acid stress on the metabolism of *C. acetobutylicum*. First, a detailed metabolic model was established through the use of ^{13}C-labeling experiments, mass spectrometry analysis, and ^{13}C-flux quantification. Metabolic fluxes were then determined under various levels of butanol and butyric acid stress and compared to no-stress condition. The results revealed robust metabolism of *C. acetobutylicum* under stress.

***Goals*:** Our goal in this project was to quantify what effects butanol and butyric acid stress have on intracellular metabolic fluxes in *C. acetobutylicum*. Measurements at the bioreactor level, including external metabolite concentrations (acetate, butyrate, butanol, ethanol, acetone), cell

growth data, and off-gas analysis revealed that overall metabolic rates are reduced under stress. However, these macroscopic measurements provide only limited information on re-distribution of intracellular metabolic fluxes. In order to quantify changes at the intracellular metabolism level in response to stress ^{13}C-labeling experiments were applied.

Methods and Results: First, we had to establish a validated model of metabolism of *C. acetobutylicum* in order to apply ^{13}C-flux analysis for flux quantification. While the biochemistry of *C. acetobutylicum* had been extensively studied in the past, central metabolic pathways remained only partially resolved. For example, several published genome-scale model provided inconsistent reconstruction of central metabolic pathways in *C. acetobutylicum*. Thus, we developed a new approach for validating metabolic network models by applying the concept of parallel ^{13}C-labeling experiments. In this approach, multiple labeling experiments are performed in parallel and the results of the parallel experiments are rigorously integrated through model-based ^{13}C-metabolic flux analysis (^{13}C-MFA) and further validated through statistical analysis. In contrast to previous qualitative studies, we used the power of integrated model-based data analysis to conclusively establish the structure of central metabolic pathways and the direction of carbon flow in the TCA cycle and through amino acid biosynthesis pathways in *C. acetobutylicum*. For this study, cells were grown on defined medium in four parallel cultures with two isotopic tracers, [1-^{13}C]glucose and [U-^{13}C]glucose, with two biological replicates for each tracer. ^{13}C-flux analysis was then applied using the Metran software (developed by Antoniewicz) to determine intracellular metabolic fluxes. For this purpose, a detailed metabolic network model of *C. acetobutylicum* metabolism was developed based on the available genome-scale models and other metabolic pathway databases. Results suggested that none of the genome-scale models could accurately reproduce experimental ^{13}C-measurements. Guided by these results and through statistical analysis, we constructed an updated network model that was able to fit all experimental data with 286 redundant measurements.

Using ^{13}C-flux analysis, we found that the TCA cycle is effectively incomplete in *C. acetobutylicum*. Specifically, there is no flux between a-ketoglutarate & succinyl-CoA, succinate & fumarate, & malate & oxaloacetate. Contrary to previously proposed hypotheses, we found that while the TCA cycle runs in the oxidative direction, the conversion of succinyl-CoA to succinate proceeds independently (**Figure 6**). As a novel finding, we identified a new growth dependent pathway in *C. acetobutylicum* that proceeds from pyruvate to fumarate. This pathway has not been observed in other organisms to our knowledge. The driving force for this pathway is the growth dependent conversion of aspartate to fumarate, as part of the biosynthesis of arginine and histidine. Our flux results also suggested that isoleucine is not produced from aspartate (or from threonine) as was previously proposed, but is instead produced exclusively from pyruvate and acetyl-CoA via the citramalate synthase pathway. Until now, no citramalate synthase gene has been reported for *C. acetobutylicum*. To identify a putative citramalate synthase gene, we performed BLASTp analysis to identify a putative citramalate synthase gene (CAC3174; currently annotated as a-isopropylmalate synthase), which is the first step in the citramalate pathway. Our flux analysis results further suggested that there may be additional metabolic cycles operating in *C. acetobutylicum* that link central carbon metabolism and amino acid metabolism. To further investigate this prediction, additional ^{13}C-labeling experiments were

performed with ^{13}C-aspartate and ^{13}C-serine tracers. We demonstrate that *C. acetobutylicum* indeed has an active metabolic cycle where carbon atoms flow from aspartate to threonine, to serine, to pyruvate, to oxaloacetate and back to aspartate. This newly identified metabolic cycle depends on two amino acid degradation reactions that are active in *C. acetobutylicum*, namely the degradation of threonine to glycine and degradation of serine to pyruvate (**Figure 6**). We believe that this cycle allows *C. acetobutylicum* to rapidly interconvert several key amino acids that are needed for cell growth. Two of the six reactions in this cycle were determined to be reversible.

With the metabolic model of *C. acetobutylicum* now firmly established, it was applied to investigate intracellular metabolism of *C. acetobutylicum* under different stress conditions: low, medium and high butanol stress (30, 60, 90 mM, respectively); and low, medium and high butyrate stress (30, 40, 50 mM, respectively). Cells were first grown on defined medium to mid-exponential growth phase (OD_{600} ~1.0). At that point, butanol or butyrate was added at the desired concentration together with a ^{13}C-tracer, either [U-^{13}C]glucose or [4-^{13}C]aspartate. Following the addition of butanol/butyrate and the tracer, cell growth was monitored by measuring optical density, overall metabolic rates were monitored by off-gas analysis (CO_2 and H_2 production rates), and intracellular metabolism was monitored by measuring the incorporation of ^{13}C-labeling into biomass using mass spectrometry. Cell growth and overall metabolic rates were significantly reduced under butyrate stress, and to a lesser extent under butanol stress (**Figure 7**). However, the ratio of H_2 production to CO_2 production, which is indicative of relative intracellular pathway utilization, was unaffected by butanol and butyrate stress (**Figure 7**). Similarly, ^{13}C-labeling incorporation into biomass was unaffected by butanol and butyrate stress (**Figure 8**).

Conclusions: Overall, these results demonstrate that *C. acetobutylicum* displays an intracellular metabolism whereby relative intracellular utilization of central pathways is maintained relatively constant at low and high levels of butanol and butyrate stress, despite a significant reduction of overall metabolic rates. This suggests that the cells employ multiple levels of regulation to maintain essential central pathways at constant relative ratios for balancing cell metabolism in health and under stress.

The Maranas group

Summary: A genome-scale metabolic (GSM) model is a powerful tool for understanding the metabolic capacities of an organism and developing metabolic-engineering strategies for strain development. The inclusion of condition-specific regulatory information to a GSM model provides additional capabilities for phenotypic predictions. We constructed and validated, and a regulated GSM model for *C. acetobutylicum* ATCC 824, *i*Cac802, which was used with experimental gene expression data to predict the cellular responses to two chemical stressors: butanol and butyrate. *i*Cac802 spans 802 genes and includes 1137 metabolites and 1462 reactions, along with gene-protein-reaction associations. Both ^{13}C-MFA and gene deletion data in the ABE fermentation pathway were used to test the predicted flux ranges allowed by the model.

Goals: The goal of this project was to understand and model the stress response of *Clostridium acetobutylicum* to two important toxic metabolites: butanol and butyrate. CoreReg method was developed to place regulation on a genome-scale metabolic model for the two stresses and identified differences in the respective responses, including distinct core sets and the restriction of biomass production similar to experimental observations. Given the core sets predicted by the CoreReg method, remedial actions can be taken to counteract the effect of stress on metabolism. The specific Aim was to develop two types but complementary models to capture essential elements of the metabolite stress responses, and build a 2nd generation Genome-Scale Model (GSM) as required for both types of modeling efforts.

Methods & Results: The 2nd generation GSM model iCac802 was constructed for *C. acetobutylicum* ATCC 824. It spans 802 genes and includes 1,137 metabolites participating in 1,462 reactions. All reactions present are elementally and charge balanced. GPR associations were determined from the available functional annotation information and homology predictions accounting for monofunctional proteins, multifunctional proteins, isozymes, and protein complexes. The model was curated to remove any thermodynamically infeasible cycles. While all previous models contained an aggregate reaction for the production of hexadecanoyl-acp and hexadecanoyl-CoA from acetyl-acp and crotonyl-CoA, respectively, iCac802 includes all participating reactions in fatty acid synthesis and metabolism pathways building up to these metabolites. *i*Cac802 also contains additional reactions from purine, pyrimidine metabolism, and cobalamin biosynthesis pathways. Both ^{13}C-MFA and gene deletion data in the ABE fermentation pathway were used to test the predicted flux ranges allowed by the model.

Regulation was incorporated into the model by a stepwise procedure that modified the minimum and maximum flux bounds of reactions based on fold change values of corresponding gene expression values by a new method called CoreReg. Gene expression data for each stress condition were used to calculate the fold change from unstressed conditions using significance analysis of microarrays (SAM). Transcriptomic data were collected for three concentrations each of butyrate and butanol. We refer to these as low (30 mM butyrate, 30 mM butanol), medium (40 mM butyrate, 60 mM butanol), and high (50 mM butyrate, 90 mM butanol) stress conditions. The fold change for each reaction under each stress condition was calculated from gene expression fold changes under stressed conditions by using GPRs. In the case of multiple enzyme subunits, the minimum expression value for the genes associated with the subunits was considered for calculating the reaction fold change. In the case of isozymes the total transcript level, obtained by summation of all isozyme transcripts, was considered. The unregulated model reaction bounds without any biomass constraint represent the minimum and maximum possible bounds of each reaction. Thus, a further increase in these bounds does not affect any maximum yield calculations, as the bounds are not active. Thus, only down-regulated genes were considered for evaluating the regulated model. The procedure for implementing regulation can be divided into five steps as explained below (see **Figure 9**).

Step 1: FVA is performed on the unregulated (*UR*) model to obtain lower $v_j^{\mathrm{L,UR}}$ and upper $v_j^{\mathrm{U,UR}}$ reaction flux bounds. The value of k is set equal to one to indicate primary core set.

Step 2: FVA upper and lower bounds for the unregulated model are multiplied by the fold change value (c_j) obtained using the transcriptomic data. The lower $v_j^{\mathrm{L,R}}$ and upper $v_j^{\mathrm{U,R}}$ bounds for the regulated (R) model are evaluated as follows:

$$\forall j \text{ such that } \mathrm{c}_j < 1 \rightarrow \begin{cases} v_j^{U,R} = v_j^{U,UR} \cdot c_j & (5) \\ v_j^{L,R} = v_j^{L,UR} \cdot c_j & (6) \end{cases}$$

Note that the updated lower bound is non-zero only for reversible reactions, effectively lowering the maximum possible flux value in the reverse direction.

Step 3: FBA is performed on the unregulated and the regulated model to obtain maximum biomass yields $v_{biom}^{\mathrm{max,UR}}$ and $v_{biom}^{\mathrm{max,R}}$, respectively. If $v_{biom}^{\mathrm{max,R}}$ varies from $v_{biom}^{\mathrm{max,UR}}$ by less than 2%, then the process is terminated, because the effect of the remaining regulation in the model is too small to cause any significant changes in metabolism as exemplified by the max biomass yield. Therefore, no additional regulatory core sets are extracted.

Step 4: FVA is performed at max biomass $v_{biom}^{\mathrm{max,R}}$ on the regulated (R) model to obtain lower $v_j^{\mathrm{l,R}}$ and upper $v_j^{\mathrm{u,R}}$ reaction flux bounds. These bounds are next compared with the imposed regulatory constraints from Step 2.

Step 5: Reactions j whose flux bounds are equal to the regulatory constraints (*that is,* $v_j^{\mathrm{u,R}} \equiv v_j^{\mathrm{U,R}}$ or $v_j^{\mathrm{l,R}} \equiv v_j^{\mathrm{L,R}}$) are assembled into the core set of reactions (of order k). To identify secondary, tertiary, and higher order core sets, the fold change values (c_j) for the previously determined k core sets are set to one, thus removing their regulatory role in the model. The process is repeated from Step 2 with the value of k increased by one.

Conclusions: The regulation placed on the model for the two stresses using CoreReg identified differences in the respective responses, including distinct core sets and the restriction of biomass production similar to experimental observations. Given transcriptomic data the CoreReg method can be used to predict an organism's response to other stressors by identifying core sets of reactions whose down-regulation propagates through stoichiometry to the remaining metabolic network causing flux changes consistent with experimentally observed trends. The CoreReg method can be applied to time varying transcriptomic data to find core sets for each time point by assuming the first time point as the basis condition. This could provide insight into the various growth phases highlighting the key changes in transcriptome. CoreReg method can also be used on the proteomic data and to analyze the proteome based on core sets and compare against experimental data.

The Wu & Huang group

***Summary*:** Omics data are integrated, whereby regulatory details of gene & protein expressions and interactions, and DNA modifications are incorporated. A network of regulons was identified to play significant roles in stress response. Moreover, methylome study unveiled motifs that are either ubiquitously or differentially modified throughout the genome. The differential DNA modifications under stress suggest a novel regulatory layer for stress orchestration.

***Goals*:** Integration of omics data to understand stress response of C. *acetobutylicum* (*Cac*).

***Methods*:** pattern-based DNA-binding motif analysis (*de novo* motif prediction with MOTIFATOR, phylogenetic footprinting), in combination with transcriptome data and comparative genomics; methylome analysis (PacBio, SMRT Portal software (v2.2.0)); protein interaction data mining; functional characterization & enrichment analysis (with FIVA software);

***Results*:** Determination of stress response network model integrating important players for the general and specialized metabolite stress response in *Cac* was achieved as shown in [1], using an exceptionally large set of temporal transcriptional data and regulon analyses. In particular, by extracting the expression patterns in transcriptome data from butanol (BuOH) & butyrate (BA) stressed cultures (through clustering via the MeV software), both common & stress specific genes are revealed. We identified 164 significantly differentially expressed transcriptional regulators and detailed the cellular programs associated with general and stressor-specific responses, many previously unexplored [1]. Pattern-based, comparative genomic analyses enabled us, for the first time, to construct a detailed picture of the genetic circuitry underlying the stress response (**Figure 10**). Notably, a list of the regulons and DNA binding motifs of the stress-related transcription factors were identified: two heat-shock response regulators, HrcA and CtsR; the SOS response regulator LexA; the redox sensor Rex; and the peroxide sensor PerR. Moreover, several transcriptional regulators controlling stress-responsive amino acid and purine metabolism and their regulons were also identified, including ArgR (arginine biosynthesis and catabolism regulator), HisR (histidine biosynthesis regulator), CymR (cysteine metabolism repressor) and PurR (purine metabolism repressor).

Based on our analysis of the single molecule real-time (SMRT) (PacBio) sequencing data of *Cac* under various culture conditions, we predicted potential DNA modification sites and DNA modification motifs. Among those motifs, three are ubiquitously methylated throughout the tested culture conditions (i.e., CTGA$^{\mathbf{m6}}$**A**G, CTTC$^{\mathbf{m6}}$**A**G and CAAAA$^{\mathbf{m6}}$**A**R). Two motifs, G$^{\mathbf{m6}}$**A**STC and G$^{\mathbf{m4}}$**C**NGC, show different level of partial differential methylation throughout the samples tested and notably under stress, under stationary phase of culture and a degenerate strain (M5) which cannot produce solvents (**Figure 11**). The differential modification of the latter two motifs led to a hypothesis of epigenetic regulation of gene expressions under stress. In particular, the differential modification of G$^{\mathbf{m6}}$**A**STC and G$^{\mathbf{m4}}$**C**NGC were not only found to be associated with certain protein-coding genes, but also some ncRNA genes (including sRNA genes [2]). Moreover, a comparison of late-stationary-phase sample versus the mid-exponential sample enabled the identification of the genes with both differential gene expression and growth-stage specific motif modifications. For the poorly understood epigenetic regulations of *Cac*, plausible regulatory loops can be suggested around the affected metabolic reactions, and the hypotheses can be tested experimentally.

We have worked with the Papoutsakis group and the Meyers group on the analysis and interpretation of the transcriptome (microarrays) and methylome (PacBio sequencing) of cultures under non-stressed, butyrate- and butanol- stressed conditions. The methylome data also include

cultures at a late-stationary phase and a mutant strain M5 (in which the megaplasmid is missing). In addition, we assisted the analysis of the RNAseq data (Illumina), by employing TPM (transcript per million) normalization for identifying high quality sequencing data.

Conclusions: Since the majority of the transcription factors and their target genes are highly conserved in other organisms of the Clostridium genus, the network discovered in our study (**Figure 10**) would be largely applicable to other Clostridium organisms. The network revealed in our study inform the molecular basis of Clostridium responses to toxic metabolites in natural ecosystems and the microbiome, and has already facilitated the construction of genome-scale models with added regulatory-network dimensions to guide the development of tolerant strains [3].

REPORT OF PRODUCTS DELIVERED:

We developed a website for the project: http://www.clostress.org/index.html. This website is accessible by everybody. Only two sub sites are password protected. These are for the monthly reports and the new data that will be first available to the project researchers.

During the 2013 Genomic Science meeting (25-27 Feb., Bethesda, MD), the Papoutsakis and Maranas groups had interactions with Dr. Chris Henry, the Microbial Science Team Co-lead of the DOE Systems Biology Knowledgebase (KBase) for incorporation of the current second generation genome scale model (GSM) of *Clostridium acetobutylicum* into KBase.

Also, the Papoutsakis and Wu & Huang groups interacted with the other DOE grantees, KBase collaborators and developers of the regulation network prediction webserver tool (RegPredict) and database (RegPrecise), Dr. Dmitri Rodionov (Sanford - Burnham Medical Research Institute) and Dr. Pavel Novichkov (Lawrence Berkeley National Laboratory). Following extensive interactions, Dr. Novichkov created an account for the Papoutsakis group to work and save the regulatory network analysis using the webserver tool, RegPredict. Also, Dr. Novichkov and Dr. Rodionov proposed to include the small regulatory RNA discovery data (consisting of 159 sRNA) in their RegPredict to facilitate robust analysis of the regulatory network in Clostridia.

PUBLICATIONS:

1. Hou S, Jones SW, Choe LH, Papoutsakis ET, Lee KH: Workflow for quantitative proteomic analysis of *Clostridium acetobutylicum* ATCC 824 using iTRAQ tags. *Methods* 2013, 61:269-276.

2. Wang Q, Venkataramanan KP, Huang H, Papoutsakis ET, Wu CH: Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress. *BMC Systems Biology* 2013, 7:120.

3. Venkataramanan, KP, Jones SW, McCormick KP, Kunjeti SG, Ralston MT, Meyers BC, Papoutsakis ET: The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in C. acetobutylicum. *BMC Genomics* 2013, 14:849.

4. Au J, Choi J, Jones SW, Venkataramanan KP, Antoniewicz MR: Parallel labeling experiments validate Clostridium acetobutylicum metabolic network model for 13C metabolic flux analysis. *Metab Eng* 2014, 26:23-33.

5. Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., & Maranas, C. D. (2014). Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. *Biotechnology for Biofuels*, 7(1), 144.

6. Venkataramanan KP, Min L, Hou S, Jones SW, Ralston MT, Lee KH, Papoutsakis ET: Integrative proteomic and transcriptomic analysis of *Clostridium acetobutylicum*'s response to

butanol and butyrate stress reveals complex post transcriptional regulation. *Biotechnology for Biofuels* 2015*, in press*, DOI: 10.1186/s13068-015-0260-9

MANUSCRIPTS IN PREPARATION:

1. Au J, Venkataramanan KP, Antoniewicz MR: Metabolic cycle between central carbon metabolism and amino acid metabolism elucidated in Clostridium acetobutylicum ATCC 824. *J Bacteriol* (In Preparation).

2. Au J, Venkataramanan KP, Antoniewicz MR: Clostridium acetobutylicum ATCC 824 displays rubust metabolism in response to butanol stress and butyric acid stress. *Metab Eng Comm* (In Preparation).

3. Au J, Antoniewicz MR: Fundamentals and applications of parallel labeling experiments for integrated 13C metabolic flux analysis: A mini-review. *Metab Eng* (In Preparation).

4. Wang Q, Venkataramanan KP, Huang H, Papoutsakis ET, Wu CH: DNA methylation in Clostridium acetobutylicum is altered under metabolite stress, in stationary phase and in the absence of the megaplasmid.

5. Ralston MT, Venkataramanan KP, Papoutsakis ET: Strand-specific RNAseq analysis of *Clostridium acetobutylicum* enables accurate transcriptome assembly revealing new genes and the transcriptional intricacies of large and small RNAs, in health and under stress.

6. Jones, A, Venkataramanan KP, Papoutsakis ET: Small, low, but potent: the role of a lowly expressed small RNA, *solB*, that regulates solvent production in *Clostridium acetobutylicum*.

POSTERS PRESENTED:

E. Terry Papoutsakis, Blake C. Meyers, Kelvin H. Lee, Maciek R. Antoniewicz, Costas D. Maranas, Hongzhan Huang & Cathy H. Wu. Systems Analysis of the *Clostridium* Metabolite stress Response for Bioenergy Applications. Presented at the: DOE Genomic Science Grantees Meeting, Bethesda, MD, Feb. 25-27, 2013.

Keerthi P. Venkataramanan, Kevin P. McCormick, Sridhara G. Kunjeti, Matthew T. Ralston, Blake C. Meyers & Eleftherios T. Papoutsakis. The small RNome of *Clostridium acetobutylicum* with and without metabolite stress. Presented at the: DOE Genomic Science Grantees Meeting, Bethesda, MD, Feb. 25-27, 2013.

Patrick F. Suthers & Costas D. Maranas. Development of a regulated model for *Clostridium acetybutilicum*. Presented at the: DOE Genomic Science Grantees Meeting, Bethesda, MD, Feb. 25-27, 2013.

Qinghua Wang, Keerthi P. Venkataramanan, Hongzhan Huang, Eleftherios T. Papoutsakis & Cathy H. Wu. Analysis of the butanol stress response regulatory network in *Clostridium*

acetobutylicum. Presented at the: DOE Genomic Science Grantees Meeting, Bethesda, MD, Feb. 25-27, 2013.

"The core regulons orchestrating the response of Clostridium acetobutylicum to butanol and butyrate stress", Qinghua Wang, Keerthi P. Venkataramanan, Hongzhan Huang, E. Terry Papoutsakis, and Cathy H. Wu, Genome Informatics 2013, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (October 30 – November 2, 2013)

"Analysis of the Butanol Stress Response Regulatory Network in Clostridium Acetobutylicum", Qinghua Wang, Hongzhan Huang, Keerthi P. Venkataramanan, E. Terry Papoutsakis, and Cathy H. Wu, Inaugural Fraunhofer – Delaware Technology Summit: Energy and Life Sciences – Solutions for Sustainability, UD, Newark, DE (March 5 and 6, 2013)

2014 ASMS Conference on Mass Spectrometry and Allied Topics, Baltimore, MD: "Proteomic and Transcriptomic Analysis of a Solvent Producing Bacterium *Clostridium acetobutylicum ATCC 824*" Min Lie, K. Venkataramanan, Hou, ET Papoutsakis, K. Lee (06/15-19/14).

Au J, Antoniewicz MR. Parallel labeling experiments: a novel approach for validating metabolic network models. Metabolic Engineering X Meeting, Vancouver (Canada). June 15, 2014

"The Core Regulons Orchestrating the Response of Clostridium acetobutylicum to Butanol and Butyrate Stress", Qinghua Wang, Keerthi P. Venkataramanan, Hongzhan Huang, E. Terry Papoutsakis, and Cathy H. Wu, Genomic Science Contractors-Grantees Meeting XII, UD, Newark, DE (February 10 - 12, 2014)

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., & Maranas, C. D. "A Regulated Model for *Clostridium acetobutylicum* Based on Response to Butanol and Butyrate Stress" Poster presented at 2014 Genomic Science Annual Contractor-Grantee Meeting; Feb 9-12; Arlington, VA.

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., & Maranas, C. D. "A Regulated Model for *Clostridium acetobutylicum* Based on Response to Butanol and Butyrate Stress" Poster presented at COBRA 2014 - 3rd Conference on Constraint-Based Reconstruction and Analysis; May 20-23; Charlottesville, VA.

Au J, Antoniewicz MR. Parallel labeling experiments: a novel approach for validating metabolic network models. 2nd Microbial Systems Symposium, Newark, DE. Feb 3, 2015

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., & Maranas, C. D. "A Regulated Model for *Clostridium acetobutylicum* Based on Response to Butanol and Butyrate Stress" Poster presented at 2015 Genomic Science Annual Contractor-Grantee Meeting; Feb 22-25; Tysons Corner, VA.

ORAL PRESENTATIONS IN CONFERENFES OR INSTITUTIONAL SEMINARS

E. Terry Papoutsakis, Shwan W. Jones, Keerthi P. Venkataramanan, Blake C. Meyers, Kevin P. McCormick, Sridhara G. Kunjeti, Kelvin H. Lee, Shuyu Hou, Maciek R. Antoniewicz, Jung

choi, Jennifer Au, Costas D. Maranas, Patrick Suthers, Qinghua Wang, Hongzhan Huang, Cathy H. Wu. TITLE: Development of a metabolite stress-response model in solventogenic clostridia by coupling multiple omic data with a genome-scale model. Session in Systems Biology: In Silico systems Biology: Cellular and Organismal Models (TA003). AIChE Annual Meeting, Pittsburg, PA, Oct 28 - Nov. 2, 2012.

Antoniewicz MR. 13C-metabolic flux analysis and parallel labeling experiments elucidate rewiring of metabolic fluxes. ECI Metabolic Engineering IX Meeting, Biarritz, France. June 3, 2012.

Patrick F. Suthers & Costas D. Maranas. A Second-Genration genome Scale Model for *Clostridium acetobutylicum*. AIChE Annual Meeting, Pittsburg, PA, Oct 28 - Nov. 2, 2012.

Keerthi P. Venkataramanan, Shawn W. Jones, Kevin P. McCormick, Sridhara G. Kunjeti, Matthew T. Ralston, Blake C. Meyers & Eleftherios T. Papoutsakis. The small Rnome of *Clostridium acetobutylicum* that responds to butanol and butyrate stress. Sesssion in: Next Generation Synthesis and Sequencing Approaches for Systems and Synthetic Biology (TA006). AIChE Annual Meeting, San Fransisco, CA, Nov 2-8, 2013.

Antoniewicz MR. Advances in 13C Metabolic Flux Analysis. Metabolic Engineering X Meeting, Vancouver (Canada). June 17, 2014.

Papoutsakis ET, Venkataramanan KP, Sandoval NR, Wang Q, Ralston MT, Juang H & Wu CH. "The methylome and deep RNome in Clostridium: new gadgets, more knowledge." CLOSTRIDIUM XIII, Shangai, China. Sep. 2014

Venkataramanan KP, Wang Q, Min L, Ralston MT, Lee KH, Wu CH, Huang H, Papoutsakis ET. "Integrated Transcriptome and Proteome Analysis in Clostridial Stress Response Model Unveils the Complexity behind the Genetic Program and Post-Transcriptional Gene Regulation." AIChE Annual Meeting, Atlanta, GA. Nov. 2014.

Sandoval NR, Venkataramanan KP, Ralston MT, Papoutsakis ET. "Characterization of a Butanol Producing Mutant of Clostridium Pasteurianum Evolved on Crude Glycerol." AIChE Annual Meeting, Atlanta, GA. Nov. 2014.

Papoutsakis ET "The methylome and deep RNome in Clostridium: new gadgets uncover larger cellular complexity" Ecole Polytechnic de Lausanne, Switzerland, Dec. 8, 2014

Antoniewicz MR. Toward a holistic understanding of cellular metabolism through 13C metabolic flux analysis. ACS BIOT Meeting, Denver, CO.

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., & Maranas, C. D. "Development of Regulated Metabolic Models for Anaerobic Organisms" Presented at 249th ACS National Meeting & Exposition, 2015; March 22-26; Denver, CO.

Au J, Antoniewicz MR. 13C metabolic flux analysis identifies a novel metabolic cycle in Clostridium acetobutylicum that involves central carbon and amino acid metabolism. SIMB Annual Meeting 2015, Philadelphia, PA. Aug 2015.

Antoniewicz MR. Advances in 13C metabolic flux analysis: parallel labeling experiments. SIMB Annual Meeting 2015, Philadelphia, PA. Aug 2015.

FIGURES for Final project report of

ID:	ER65257-1038489-0017559
Principal Investigator:	Eleftherios Papoutsakis 302-831-8376
Co-PIs:	Blake C. Meyers, Kelvin H. Lee, Maciek R. Antoniewicz, Costas D. Maranas, Hongzhan Huang and Cathy H. Wu.
Institution:	University of Delaware
Title:	Experimental Systems-Biology Approaches for Clostridia-Based Bioenergy Production
SC Division:	SC-23.2
Program Manager:	Dean A. Cole; Roland Hirsch (BER; Ph: 301-903-9009; Roland.Hirsch@science.doe.gov)

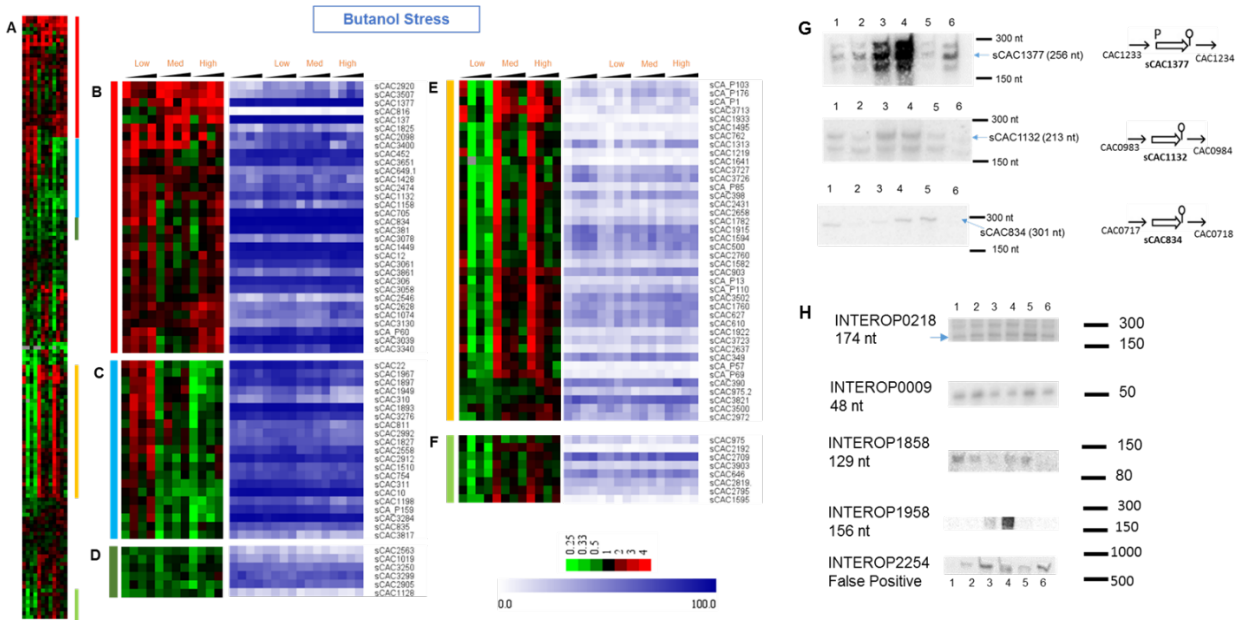


Figure 1. Analysis of the expression of small regulatory non-coding RNA (sRNA) in *C. acetobutylicum* under metabolite stress using RNA deep sequencing. A to F represents the hierarchical clustering of the 159 sRNAs in *C. acetobutylicum* under butanol stress. The 159 sRNAs consists of two subsets: 113 previously predicted & 46 newly identified (Papoutsakis & Meyers group) using RNA-seq data. 124 of these sRNAs were found to be differentially expressed under stress. Northern analysis of the selected sRNAs from (G) 113 previously predicted sRNAs & (H) 46 newly identified sRNAs.

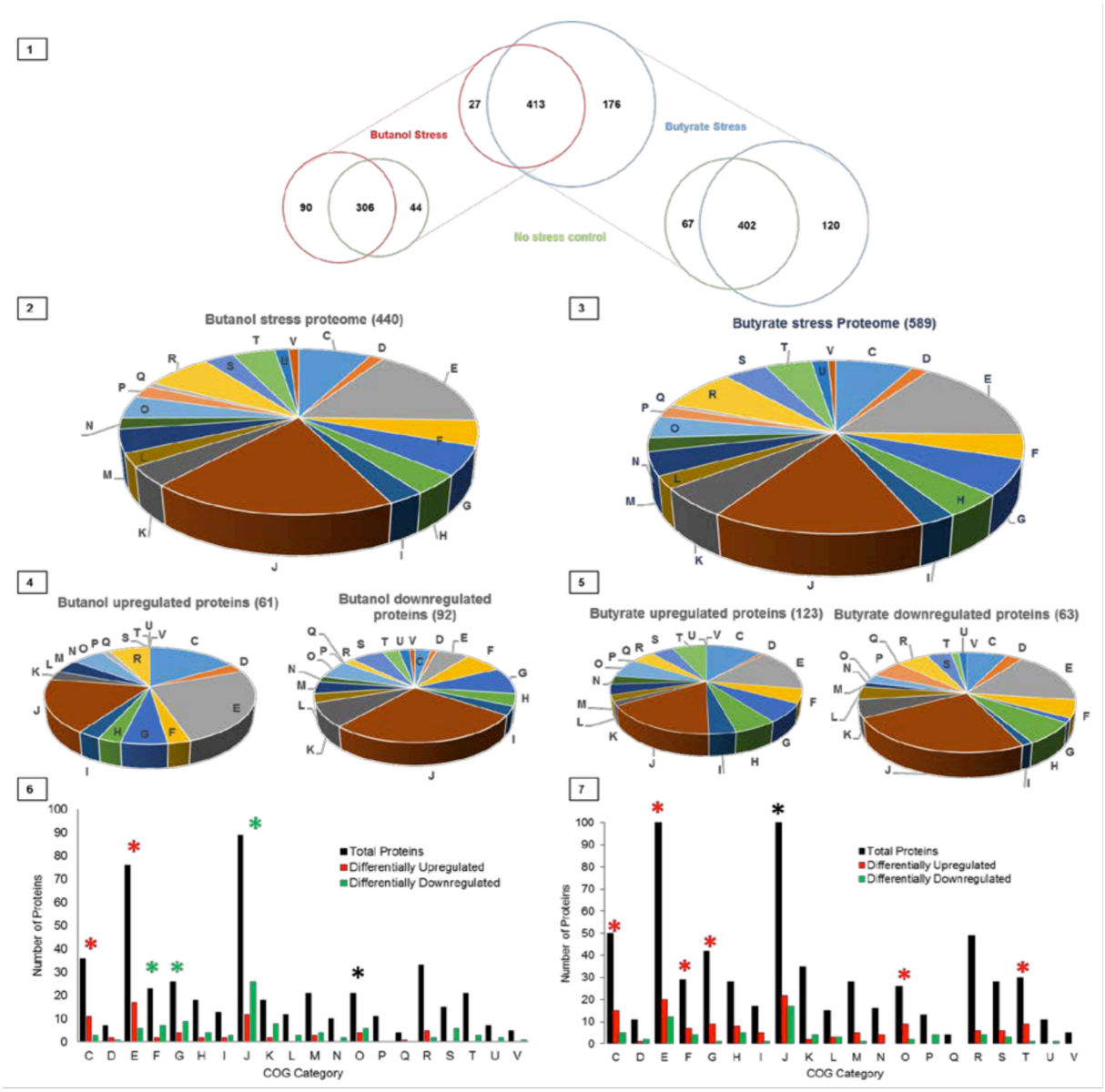


Figure 2. Clostridial proteomic summary under metabolite stress. (Panel 1) Comparison of the butanol (red) and butyrate (blue) stress proteome. Comparison of proteome between non-stress control condition and (Panel 2) butanol stress. Comparison of proteome between non-stress control condition and (Panel 3) butyrate stress. Distribution of the stress proteome into various COG functional groups (Panel 4) butanol stress and (Panel 5) butyrate stress. Differential expression within COG categories (Panel 6) butanol stress and (Panel 7) butyrate stress * - OCG gcategory enriched with upregulated proteins; * - COG category enriched in downregulated proetins; * - COG category equally enriched in up- and down-regulated proteins. C: Energy production and conversion; D: Cell division and chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme metabolism; I: Lipid metabolism; J: Translation, ribosomal

structure and biogenesis; K: Transcription; L: DNA replication, recombination and repair; M: Cell envelope biogenesis, outer membrane; N: Cell motility and secretion; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms.

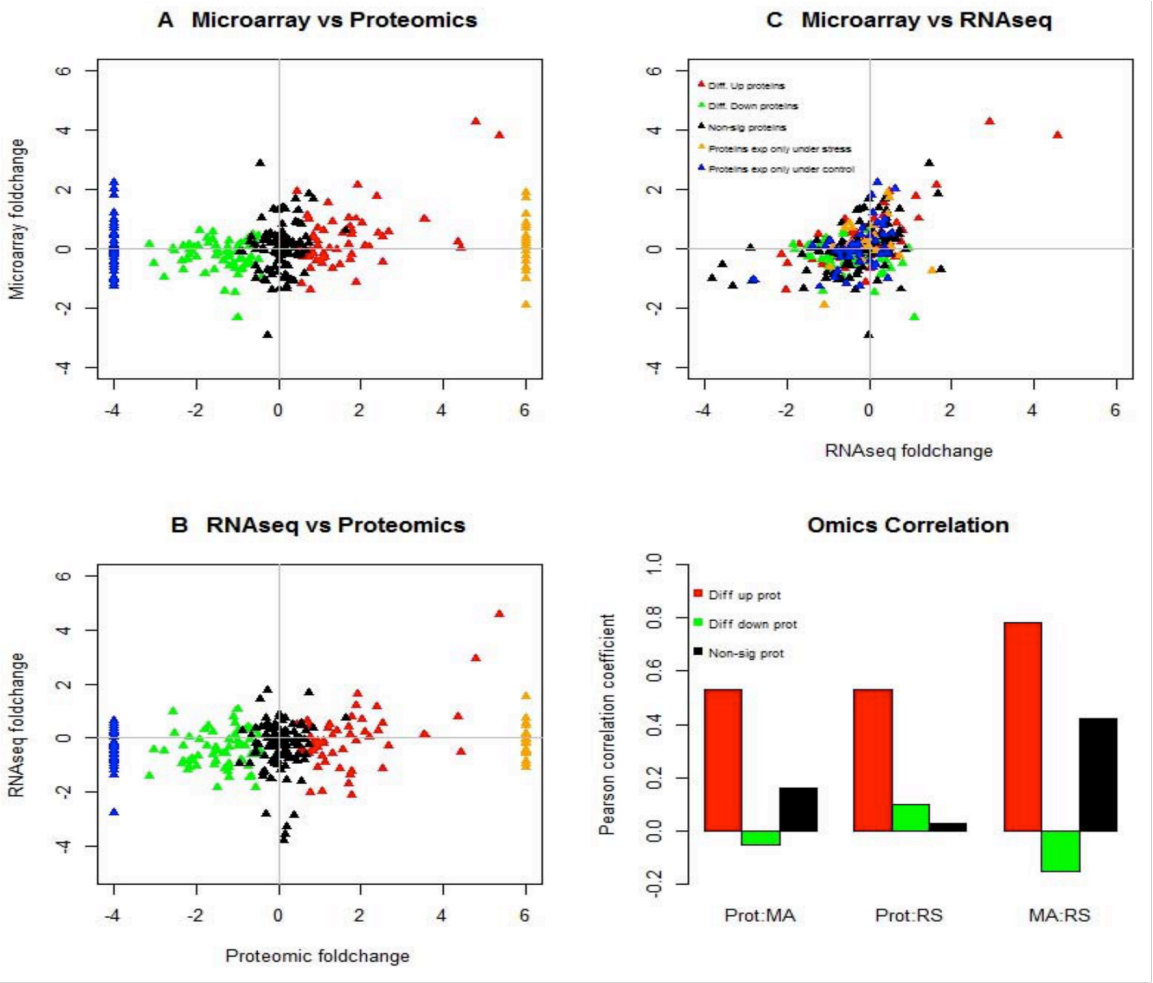


Figure 3. Comparison and correlation between proteomic and transcriptomic data under high butanol stress. A) Microarray vs proteomic comparison. B) RNAseq vs proteomic comparison. C) Microarray vs RNAseq comparison. D) Pearson correlation. All significant expression are with respect to proteomic data only. Red – differentially upregulated proteins; Green – differentially downregulated proteins; Black – non-significant proteins; Blue – proteins expressed only under non-stress control; Orange – proteins expressed only under stress. Genes/proteins lacking expression were represented by gray color.

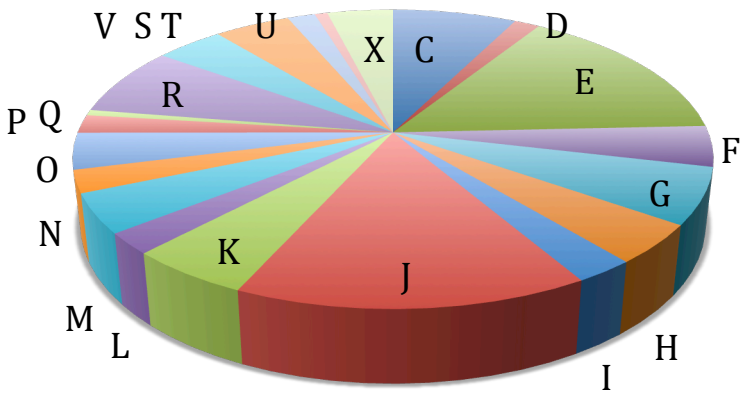


Figure 4. The distribution of COG categories for identified proteins in *C. acetobutylicum* ATCC 824 upon butyrate stress. Annotations: J, Translation; K, Transcription; L, Replication, recombination & repair; D, Cell division & chromosome partitioning; O, Posttranslational modification, protein turnover, chaperones; M, Cell envelope biogenesis, outer membrane; N, Cell motility & secretion; P, Inorganic ion transport & metabolism; T, Signal transduction mechanisms; C, Energy production & conversion; G, Carbohydrate transport & metabolism; E, Amino acid transport & metabolism; F, Nucleotide transport & metabolism; H, Coenzyme metabolism; I, Lipid metabolism; Q, Secondary metabolites biosynthesis, transport & catabolism; R, General function prediction only; S, Function unknown; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms;
X, Not in COGs.

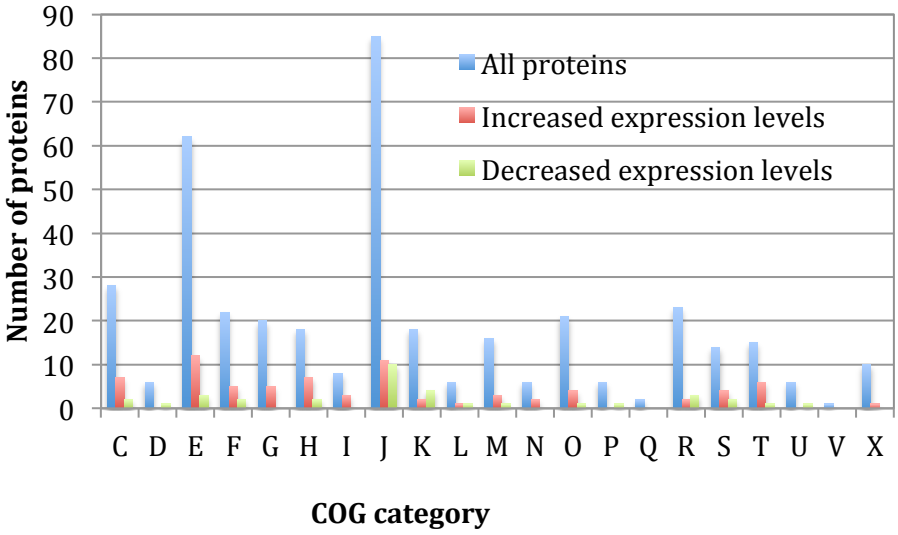


Figure 5. The distribution of COG categories for identified proteins and proteins with differential expressions in *Cac* under butyrate stress.

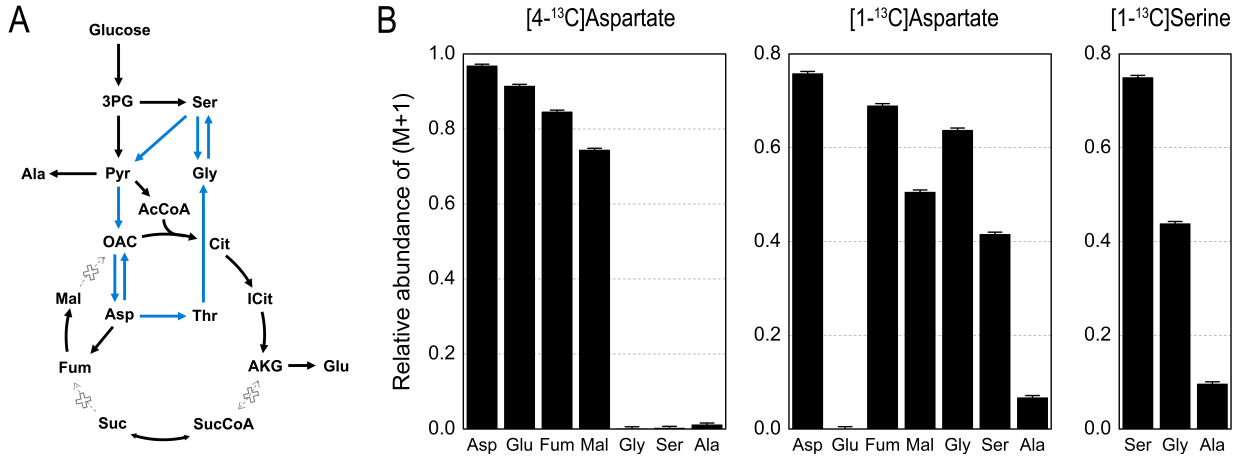


Figure 6. (A) Schematic of the metabolism of *Clostridium acetobutylicum*, including the metabolic cycle between central carbon metabolism and amino acid metabolism identified in this study (highlighted with blue arrows). (B) Relative abundances of M+1 mass isotopomer in intracellular metabolites from tracer experiments with [4-^{13}C]aspartate, [1-^{13}C]aspartate and [1-^{13}C]serine. The ^{13}C-labeling data validated the presence of the metabolic cycle and the reversibility of the reactions between aspartate and oxaloacetate, and serine and glycine.

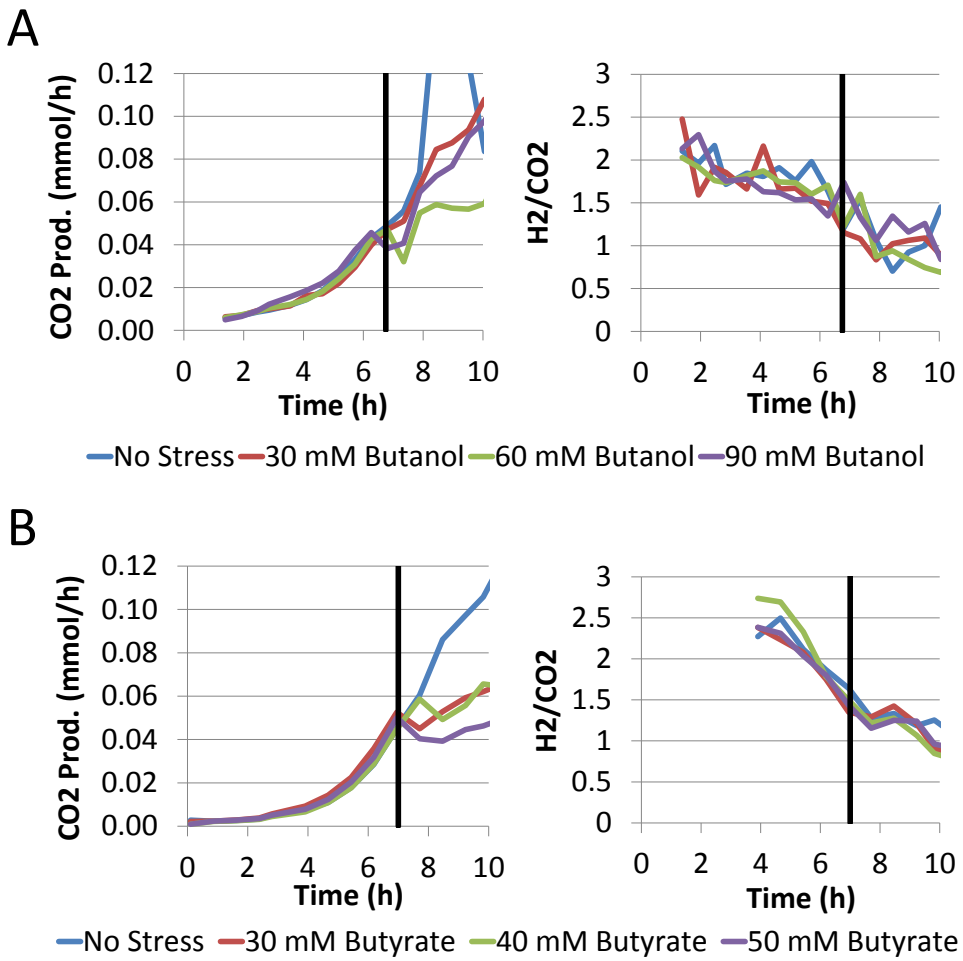


Figure 7. Off-gas analysis of *Clostridium acetobutylicum* cultures under different levels of butanol stress (A) and butyrate stress (B). The black vertical lines indicate the time point when butanol or butyrate was added to the culture. The absolute metabolic rate of CO_2 production was reduced in response to butanol and butyrate stress, however, the ratio of hydrogen production to CO_2 production was not affected by butanol or butyrate stress.

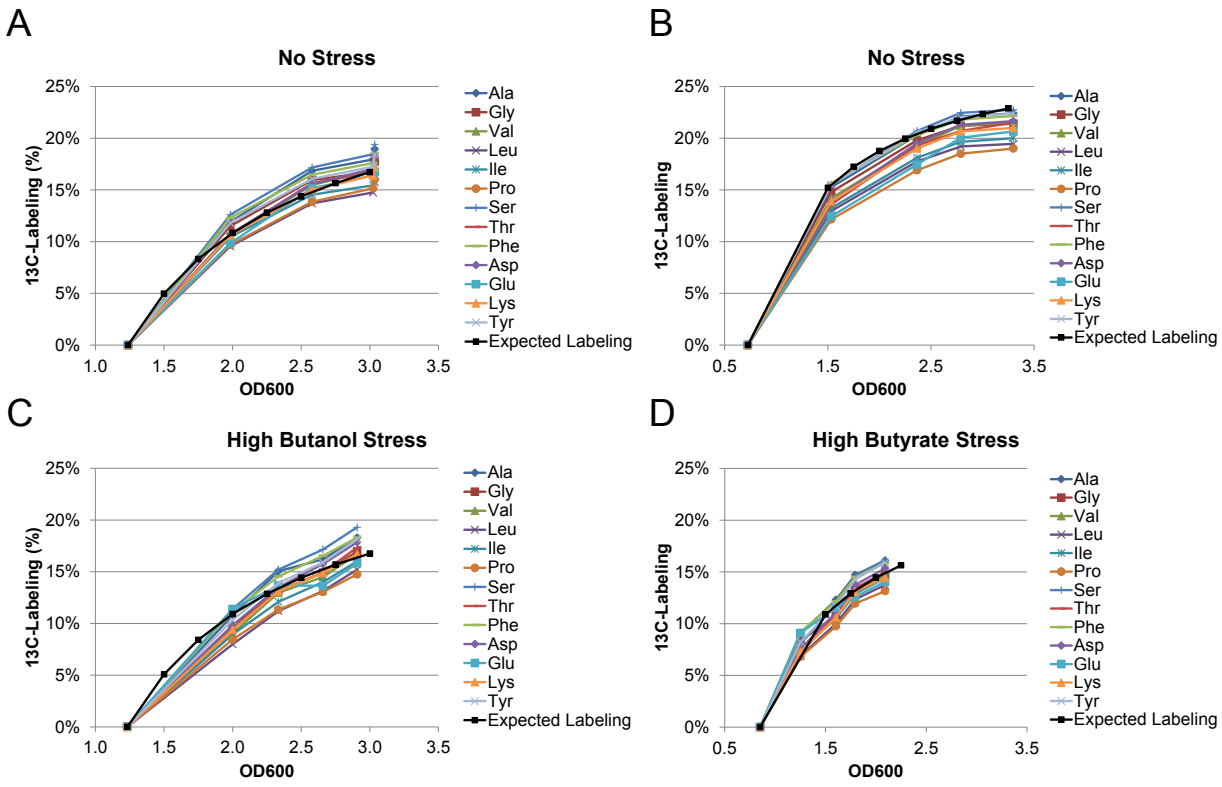


Figure 8. Incorporation of ^{13}C-labeling into biomass amino acids under high butanol stress (C) and high butyrate stress (D) compared to respective no-stress controls (A and B). The time profiles of ^{13}C-labeling incorporation followed the expected labeling trends (black lines) assuming no changes in intracellular metabolic fluxes. Thus, these results suggest that relative metabolic fluxes are not affected by butanol and butyrate stress.

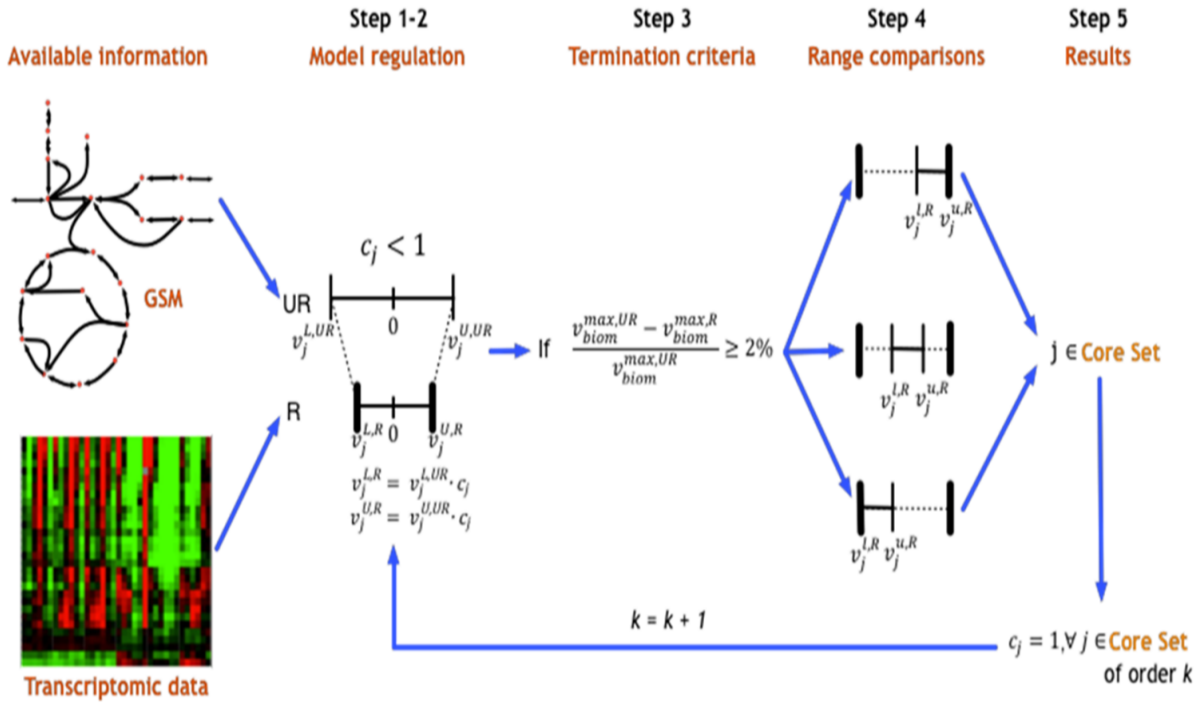


Figure 9: Graphical representation of the CoreReg procedure

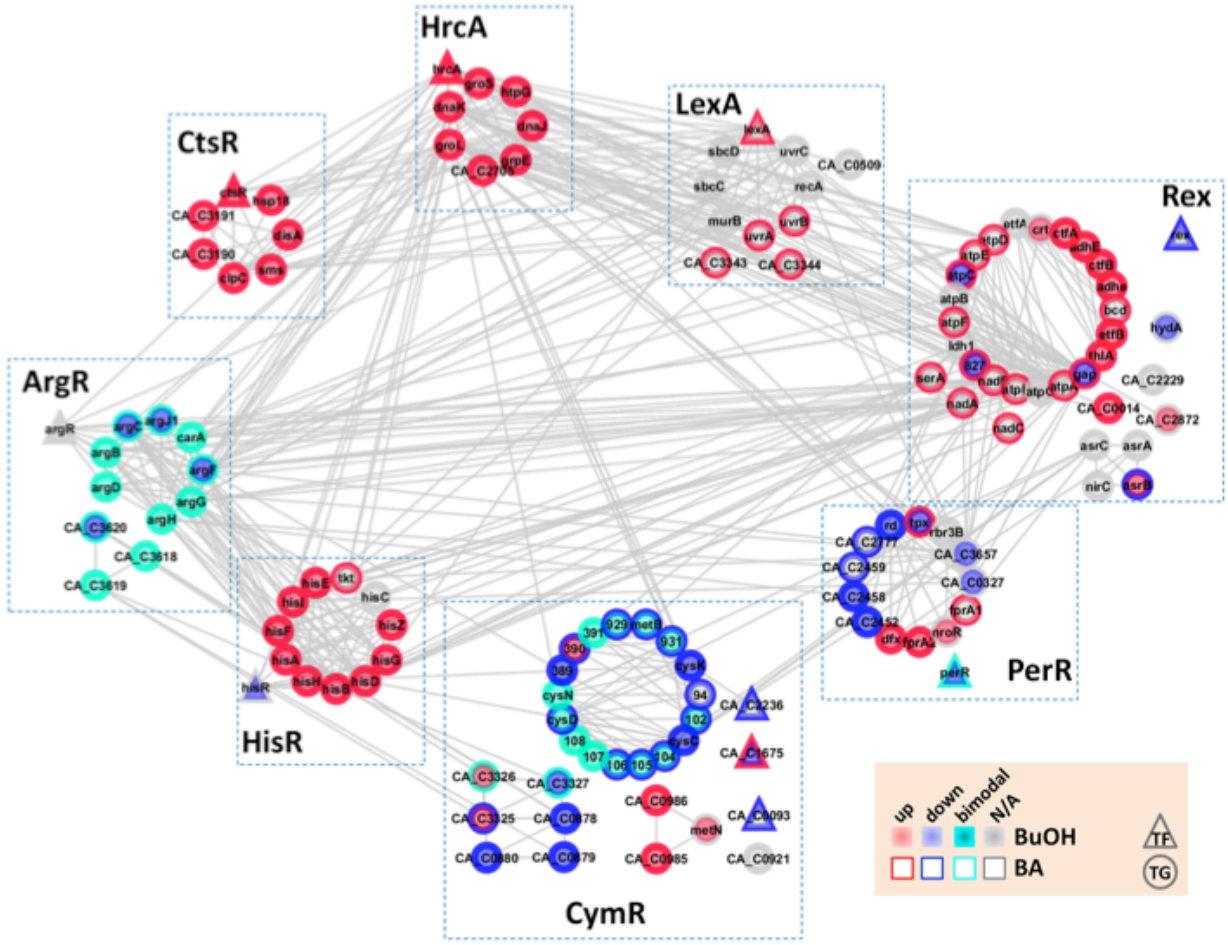
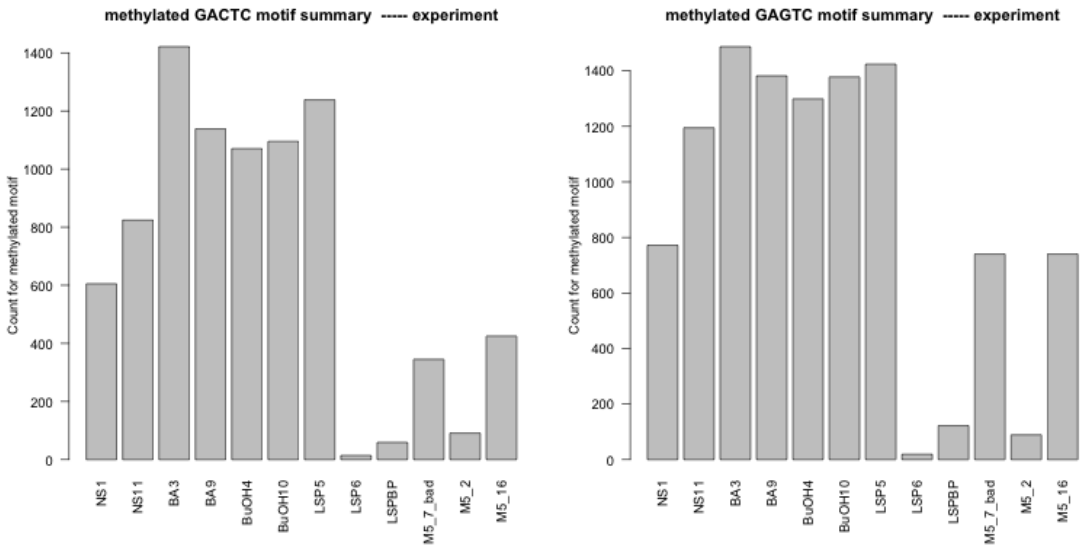
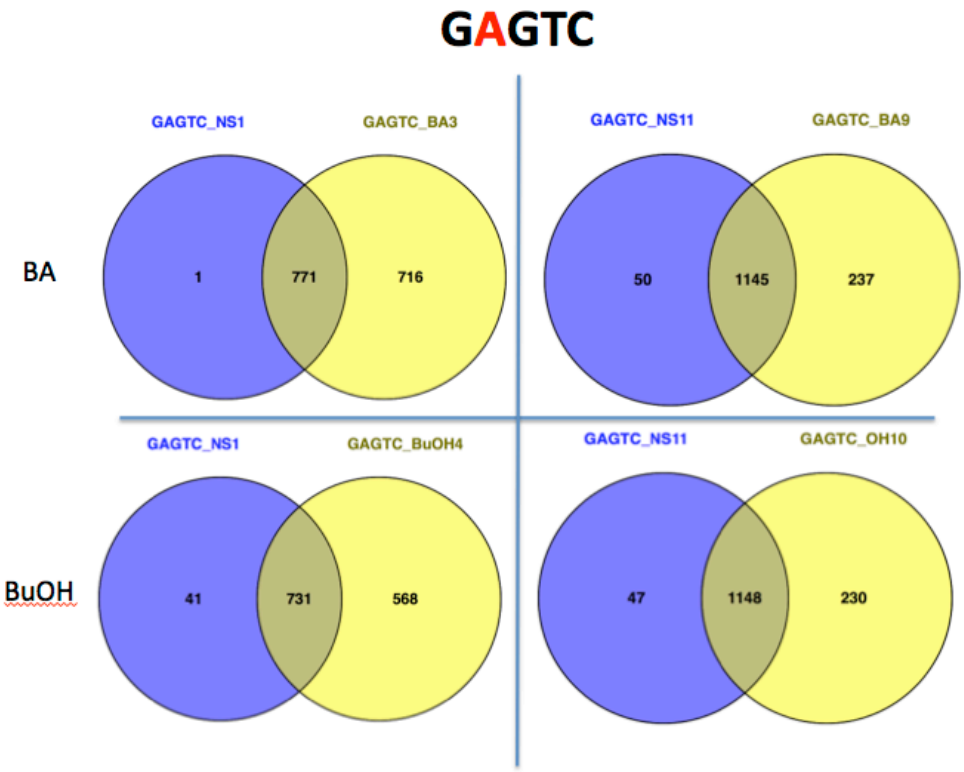


Figure 10. Stress response regulations. On the top, the STRING-based regulatory protein interaction network with 8 key regulons identified in this study. Transcription factors are in triangle, and target genes in circle. Filled color or edge color corresponds to up-, down-, bimodal- gene expression under butanol or butyrate stress. The edges connecting nodes are derived from STRING database. The differential expression of stress-responsive sRNA genes and differential modification of DNA motifs can integrate with the above regulon network to synergistically orchestrate stress response.

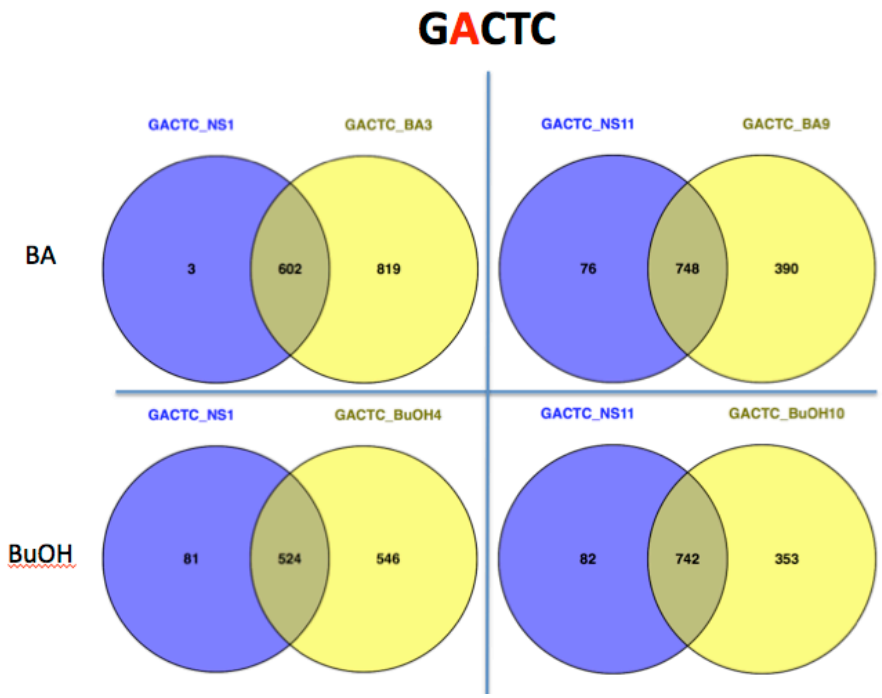
A


methylated GACTC motif summary ----- experiment
Count for methylated motif
0
200
400
600
800
1000
1200
1400
NS1
NS11
BA3
BA9
BuOH4
BuOH10
LSP5
LSP6
LSPBP
M5_7_bad
M5_2
M5_16
methylated GAGTC motif summary ----- experiment
Count for methylated motif
0
200
400
600
800
1000
1200
1400
NS1
NS11
BA3
BA9
BuOH4
BuOH10
LSP5
LSP6
LSPBP
M5_7_bad
M5_2
M5_16

B


GAGTC
GAGTC_NS1
GAGTC_BA3
GAGTC_NS11
GAGTC_BA9
BA
1
771
716
50
1145
237
GAGTC_NS1
GAGTC_BuOH4
GAGTC_NS11
GAGTC_OH10
BuOH
41
731
568
47
1148
230

C



D

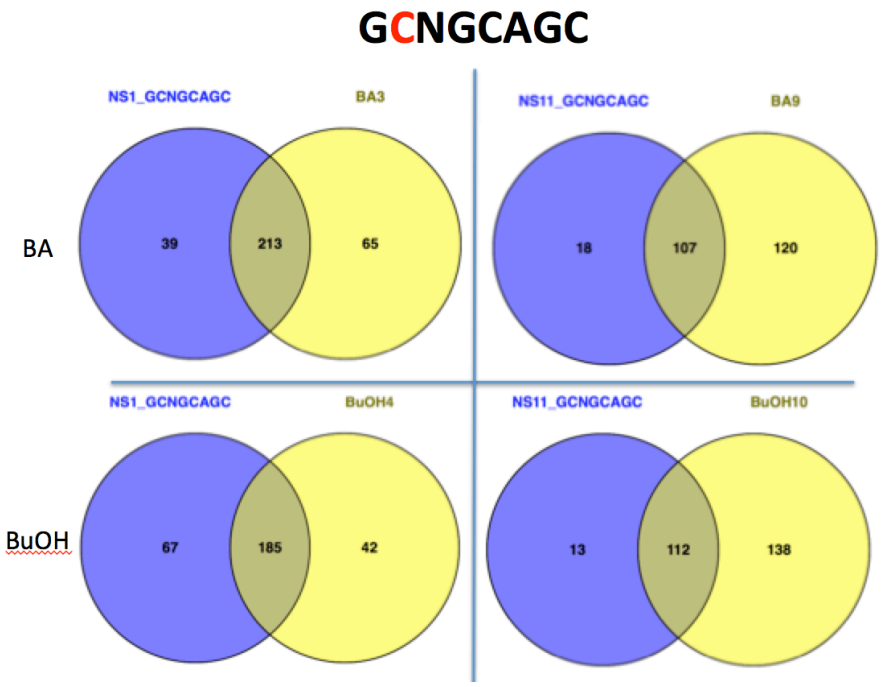


Figure 11. (A) methylation of GASTC under different culture conditions. (B-D) Venn diagrams for differentially methylated G**A**GTC (B), G**A**CTC (C), and G**C**NGCAGC (D) sites under butyrate and butanol stresses, with non-stressed sample as control. Results from wo biological replicates are shown.