A Quantitative System-Scale Characterization of the Metabolism of Clostridium acetobutylicum

Minyong Yoo,a,b,c Gwenaelle Bestel-Corre,d Christian Croux,a,b,c Antoine Riviere,a,b,c Isabelle Meynial-Salles,a,b,c Philippe Soucaille,a,b,c,d

Université de Toulouse, INSA, UPS, INP, LISBP, Toulouse, France; INRA, UMR792, Toulouse, France; CNRS, UMR5504, Toulouse, France; Metabolic Explorer, Biopôle Clermont-Limagne, Saint Beauzire, France

ABSTRACT Engineering industrial microorganisms for ambitious applications, for example, the production of second-generation biofuels such as butanol, is impeded by a lack of knowledge of primary metabolism and its regulation. A quantitative system-scale analysis was applied to the biofuel-producing bacterium Clostridium acetobutylicum, a microorganism used for the industrial production of solvent. An improved genome-scale model, iCac967, was first developed based on thorough biochemical characterizations of 15 key metabolic enzymes and on extensive literature analysis to acquire accurate fluxomic data. In parallel, quantitative transcriptomic and proteomic analyses were performed to assess the number of mRNA molecules per cell for all genes under acidogenic, solventogenic, and alcohologenic steady-state conditions as well as the number of cytosolic protein molecules per cell for approximately 700 genes under at least one of the three steady-state conditions. A complete fluxomic, transcriptomic, and proteomic analysis applied to different metabolic states allowed us to better understand the regulation of primary metabolism. Moreover, this analysis enabled the functional characterization of numerous enzymes involved in primary metabolism, including (i) the enzymes involved in the two different butanol pathways and their cofactor specificities, (ii) the primary hydrogenase and its redox partner, (iii) the major butyryl coenzyme A (butyryl-CoA) dehydrogenase, and (iv) the major glyceraldehyde-3-phosphate dehydrogenase. This study provides important information for further metabolic engineering of C. acetobutylicum to develop a commercial process for the production of n-butanol.

IMPORTANCE Currently, there is a resurgence of interest in Clostridium acetobutylicum, the biocatalyst of the historical Weizmann process, to produce n-butanol for use both as a bulk chemical and as a renewable alternative transportation fuel. To develop a commercial process for the production of n-butanol via a metabolic engineering approach, it is necessary to better characterize both the primary metabolism of C. acetobutylicum and its regulation. Here, we apply a quantitative system-scale analysis to acidogenic, solventogenic, and alcohologenic steady-state C. acetobutylicum cells and report for the first time quantitative transcriptomic, proteomic, and fluxomic data. This approach allows for a better understanding of the regulation of primary metabolism and for the functional characterization of numerous enzymes involved in primary metabolism.

Received 20 October 2015 Accepted 26 October 2015 Published 24 November 2015

C. acetobutylicum is a Gram-positive, spore-forming anaerobic bacterium capable of converting various sugars and polysaccharides to organic acids (acetate and butyrate) and solvents (acetone, butanol, and ethanol). Due to its importance in the industrial production of the bulk chemicals acetone and butanol (1–3) and its potential use in the production of n-butanol, a promising biotechnology-based liquid fuel with several advantages over ethanol (4, 5), much research has focused on understanding (i) the regulation of solvent formation (6–13), (ii) the ability to tolerate butanol (14–17), and (iii) the molecular mechanism of strain degeneration in C. acetobutylicum (18, 19). The complete genome sequence of C. acetobutylicum ATCC 824 has been published (20), and numerous transcriptomic and proteomic studies have been performed to date (21–26). Although most of these transcriptomic studies have been performed using two-color microarrays (25, 27–29), RNA deep sequencing (RNA-seq) has recently been used, allowing a more accurate quantification of transcripts as well as the determination of transcription start sites and 5’ untranscribed regions (5’ UTRs) (17, 30). With regard to proteomic studies of C. acetobutylicum, 2-dimensional gel electrophoresis (2-DGE) (22, 24, 31–33) is typically employed. 2-DGE is popular and generates substantially valuable data; however, limitations of this method, such as low reproducibility, narrow dynamic range, and low throughput, remain (34). Recently, more quantitative approaches have been developed using two-dimensional liquid chromatography–tandem mass spectrometry (2D-LC-MS/MS) (35) or iTRAQ tags (36).

In general, transcriptomic and/or proteomic studies of C. acetobutylicum have been focused on understanding (i) the transcriptional program underlying spore formation (21, 23), (ii) the re-
sponse or adaptation to butanol and butyrate stress (14–17), and (iii) the regulation of primary metabolism (21–23, 25, 35, 37).

Furthermore, to elucidate the molecular mechanisms of endospore formation, microarrays (21, 23) have been used extensively in combination with the downregulation of sigma factors by antisense RNA (23) or inactivation by gene knockout (38, 39). Initially, investigations of the response of \( C. \text{acetobutylicum} \) to butanol and butyrate stress employed microarrays (14–16) followed by RNA deep sequencing (RNA-seq) to quantify both mRNA and small noncoding RNAs (sRNA) (17), and quantitative transcriptomic and proteomic approaches were later combined (40). Based on one of these studies (16), regulons and DNA-binding motifs of stress-related transcription factors as well as transcriptional regulators controlling stress-responsive amino acid and purine metabolism and their regulons have been identified. Furthermore, integrative proteomic-transcriptomic analysis has revealed the complex expression patterns of a large fraction of the proteome that could be explained only by involving specific molecular mechanisms of posttranscriptional regulation (40).

The regulation of solvent formation in \( C. \text{acetobutylicum} \) has been extensively studied by microarrays (21, 23, 25) and/or proteomic (24, 35) approaches. Despite the valuable insights achieved in those studies, many physiological parameters, such as specific growth rates, specific glucose consumption rates, pH, and cellular differentiation, as well as butyrate and butanol stress, change with time, making it difficult to understand many details of the expression pattern.

In phosphate-limited chemostat cultures, \( C. \text{acetobutylicum} \) can be maintained in three different stable metabolic states (6, 8–10, 41) without cellular differentiation (37): aci-dogenic (producing acetate and butyrate) when grown at neutral pH on glucose, solventogenic (producing acetone, butanol, and ethanol) when grown at low pH on glucose, and alcohologenic (forming butanol and ethanol but not acetone) when grown at neutral pH under conditions of high NAD(P)H availability. Indeed, because the cells are maintained under steady-state conditions with constant endogenous and exogenous parameters such as a specific growth rate and specific substrate consumption rate, chemostat culture is the preferred fermentation method by which to achieve standardized conditions with a maximum degree of reproducibility. Transcriptional analysis of the transition from an aci-dogenic to a solventogenic state (37), as well as transcriptomic and proteomic analyses of acidicogenic and solventogenic (22) phosphate-limited chemostat cultures, has already been performed using two-color microarrays for transcriptomic analysis and 2-DGE for proteomic analysis, methods that are semiquantitative. However, a systems biology approach, combining more than two quantitative “omic” analyses of chemostat cultures of \( C. \text{acetobutylicum} \), has never been performed.

Therefore, the aim of this study was to apply a quantitative system-scale analysis to acidicogenic, solventogenic, and alcohologenic steady-state \( C. \text{acetobutylicum} \) cells to provide new insight into the metabolism of this bacterium. We first developed an improved genome-scale model (GSM), including a thorough biochemical characterization of 15 key metabolic enzymes, to obtain accurate fluxomic data. We then applied quantitative transcriptomic and proteomic approaches to better characterize the distribution of carbon and electron fluxes under different physiological conditions and the regulation of \( C. \text{acetobutylicum} \) metabolism.

### RESULTS AND DISCUSSION

**Improving upon current GSMs for metabolic flux analysis.** The \( iCac967 \) model for \( C. \text{acetobutylicum} \) ATCC 824 spans 967 genes and includes 1,058 metabolites participating in 1,231 reactions (Table 1; also see Data Set S1 in the supplemental material). All reactions are elementally and charge balanced. The \( iCac967 \) model is the result of an extensive literature analysis associated with the biochemical characterization of many key metabolic enzymes in an attempt to better understand the distribution of carbon and electron fluxes. The previously uncharacterized butyryl-CoA dehydrogenase (42) was demonstrated that the butyryl-CoA dehydrogenase of \( C. \text{acetobutylicum} \) is a strictly NADH-dependent enzyme and that ferredoxin is needed for the reaction to proceed. To study the stoichiometry of the reaction, the concentrations of NADH (see Fig. S1A) and crotonyl-CoA (see Fig. S1B) were modulated using constant concentrations of purified ferredoxin (CA_C0303) and hydrogenase (CA_C0028). Based on the initial slope in Fig. S1B in the supplemental material, it was calculated that in the presence of excess crotonyl-CoA, 2.15 mol of NADH was required for the formation of 1 mol of \( H_2 \); from the initial slope in Fig. S1A in the supplemental material, it was calculated that in the presence of excess NADH, 1.25 mol of crotonyl-CoA was required for the formation of 1 mol of \( H_2 \). The results indicate that under fully coupled conditions, approximately 1 mol of ferredoxin is reduced by 2 mol of NADH and 1 mol of crotonyl-CoA, similar to the butyryl-CoA dehydrogenase of \( Clostridium \text{kluyveri} \) (43). Although the possibility that this enzyme might consume 2 mol of NADH and produce 1 mol of reduced ferredoxin in \( C. \text{acetobutylicum} \) was previously presented as a hypothesis (44), it has not been demonstrated to date, nor has it been integrated in the recently published GSMs (45, 46). This result has strong implications for the distribution of electron fluxes, as discussed below in the metabolic flux analysis section.

| TABLE 1 Comparison of GSMs of \( C. \text{acetobutylicum} \)* |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Model statistic                 | No. of genes, reactions, or metabolites in GSM: |
| Senger et al. (56, 57)          | 474 432 490     | 802 967         |
| Lee et al. (58)                 | 552 502 794     | 1,462 1,231     |
| McAnulty et al. (46)            | 422 479 707     | 1,137 1,058     |
| Dash et al. (45)                |                 |                 |
| *Cac967*                        |                 |                 |

*The numbers of genes, reactions, and metabolites present in four previous GSMs of \( C. \text{acetobutylicum} \) and \( iCac967 \) are shown.*
contrast, AdhE2 possesses both high butyraldehyde and butanol dehydrogenase activities with NADH. The three potential alcohol dehydrogenases, BdhA, BdhB, and BdhC (49), encoded by bdhA, bdhB, and bdhC (CA_C3299, CA_C3298, and CA_C3392), respectively, were heterologously expressed in *E. coli* and then characterized after purification as tag-free proteins (Table 2). The three enzymes were demonstrated to be primarily NADPH-dependent butanol dehydrogenases, results which do not agree with the previous characterizations of BDH and BDHII (later demonstrated to be encoded by bdhA and bdhB), which were reported to be NADH dependent (49, 50). However, in agreement with our findings, all of the key amino acids of the two GGGS motifs at positions 37 to 40 and 93 to 96 involved in the NADPH binding of YqhD, a strictly NADPH-dependent alcohol dehydrogenase (51), are perfectly conserved in the three *C. acetobutylicum* alcohol dehydrogenases. Furthermore, these results are also in line with previously published data from two different research groups (9, 52) showing that in a crude extract of solventogenic *C. acetobutylicum* cultures, the butanol dehydrogenase activity measured in the physiological direction is mainly NADPH dependent. As discussed below, *C. acetobutylicum* must utilize at least one of these alcohol dehydrogenases to produce butanol and ethanol under solventogenic conditions, which implies that 1 mol of NADPH is needed for each mole of butanol and ethanol produced under solventogenic conditions.

The cofactor specificity of the ammonium assimilation pathway that proceeds via glutamine 2-oxoglutarate aminotransferase (GOGAT) encoded by gltA and gltB (CA_C1673 and CA_C1674, respectively) and glutamate dehydrogenase (GDH) encoded by gdh (CA_C0737) was also characterized. The gltA-gltB and gdh genes were expressed in *C. acetobutylicum* and *E. coli*, respectively, and GOGAT and GDH were purified (Table 2). Both enzymes were found to be NADH dependent, in contrast to the corresponding enzymes in *E. coli*, which are NADPH dependent (53, 54).

The functions of the three genes (CA_C0970, CA_C0971, and CA_C0972) proposed (55) to encode the first three steps of the oxidative branch of the tricarboxylic acid (TCA) cycle were unambiguously characterized. CA_C0970, CA_C0971, and CA_C0972 were individually expressed in *E. coli*, and their gene products were purified (Table 2); the genes were demonstrated to encode an *Re*-citrate synthase (CitA), an aconitase (CitB), and an NADH-dependent isocitrate dehydrogenase (CitC), respectively.

Finally, we characterized the cofactor specificity of the two malic enzymes encoded by CA_C1589 and CA_C1596, almost-identical genes that differ by only two nucleotides. Not surprisingly, the specific activities of the two purified enzymes are almost identical, and both enzymes are NADPH dependent (Table 2).

The iCac967 model statistics and those of all other published models for *C. acetobutylicum* (45, 46, 56–58) are shown in Table 1. iCac967 has 20% more genes than the most recently published model by Dash et al. (45) but fewer metabolites and reactions, as some reactions described by these authors were not validated by our extensive literature analysis or were inappropriate in the context of anaerobic metabolism, for example, R0013 (NADPH + O₂ + H⁺ + 2-octaprenylphenol → H₂O + NADP⁺ + 2-octaprenyl-6-hydroxyphenol) and R0293 (H₂O + O₂ + sarcosine → H₂O₂ + glycine + formaldehyde). Furthermore, we applied our DSM to the butyrate kinase knockout mutant (59) and the M5 degenerate strain (60) (which has lost the pSOL1 plasmid) and successfully predicted their phenotypes (see Table S1 in the supplemental material).

**Quantitative transcriptomic and proteomic analyses of *C. acetobutylicum* under stable acidogenic, solventogenic, and alcohologenic conditions.** (i) General considerations. Quantitative transcriptomic and proteomic analyses were performed on phosphate-limited chemostat cultures of *C. acetobutylicum* maintained in three different stable metabolic states: acidogenic, solventogenic, and alcohologenic (6, 7, 9, 10). The total amount of DNA, RNA, and protein contents (expressed in grams per gram of dry cell weight [DCW]) and the number of cells per gram of DCW were experimentally determined for each steady-state condition under phosphate limitation at a dilution rate of 0.05 h⁻¹. These numbers were not significantly different among the steady-state conditions, in agreement with previous studies (61, 62) on *E. coli* that have shown that the biomass composition is not dependent on the carbon source but is strictly dependent on the specific growth rate. According to all of the values, the average contents of DNA (1.92 ± 0.03), mRNA ([9.41 ± 0.94] × 10⁵), and protein

---

**TABLE 2 Activities of purified key metabolic enzymes**

<table>
<thead>
<tr>
<th>Locus no.</th>
<th>Gene name</th>
<th>Enzyme activity</th>
<th>Activity (U/mg)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA_C3299</td>
<td>bdhA</td>
<td>Butanol dehydrogenase</td>
<td>NADH (0.15 ± 0.05), NADPH (2.57 ± 0.45)</td>
</tr>
<tr>
<td>CA_C3298</td>
<td>bdhB</td>
<td>Butanol dehydrogenase</td>
<td>NADH (0.18 ± 0.02), NADPH (2.95 ± 0.36)</td>
</tr>
<tr>
<td>CA_C3392</td>
<td>bdhC</td>
<td>Butanol dehydrogenase</td>
<td>NADH (0.24 ± 0.04), NADPH (2.21 ± 0.41)</td>
</tr>
<tr>
<td>CA_P0035</td>
<td>adhE1</td>
<td>Butanol dehydrogenase</td>
<td>NADH (0.04 ± 0.02), NADPH (not detected)</td>
</tr>
<tr>
<td>CA_P0162</td>
<td>adhE2</td>
<td>Butyaldehyde dehydrogenase</td>
<td>NADH (4.8 ± 0.42), NADPH (0.12 ± 0.01)</td>
</tr>
<tr>
<td>CA_P0162</td>
<td>adhE1</td>
<td>Butyaldehyde dehydrogenase</td>
<td>NADH (2.27 ± 0.21), NADPH (0.08 ± 0.01)</td>
</tr>
<tr>
<td>CA_P0035</td>
<td>adhE2</td>
<td>Butyaldehyde dehydrogenase</td>
<td>NADH (2.5 ± 0.31), NADPH (0.07 ± 0.01)</td>
</tr>
<tr>
<td>CA_C2711-CA_C2709</td>
<td>bcd-etfB-etfA</td>
<td>Butyryl-CoA dehydrogenase</td>
<td>NADH (0.569 ± 0.08), NADPH (not detected)</td>
</tr>
<tr>
<td>CA_C1673-CA_C1674</td>
<td>gltA/gltB</td>
<td>Glutamate synthase</td>
<td>NADH (0.61 ± 0.16), NADPH (0.051 ± 0.01)</td>
</tr>
<tr>
<td>CA_C0737</td>
<td>gdh</td>
<td>Glutamate dehydrogenase</td>
<td>NADH (41.2 ± 3.4), NADPH (0.12 ± 0.01)</td>
</tr>
<tr>
<td>CA_C0970</td>
<td>citA</td>
<td>Re-citrate synthase</td>
<td>1.9 ± 0.14</td>
</tr>
<tr>
<td>CA_C0971</td>
<td>citB</td>
<td>Aconitase</td>
<td>6.5 ± 0.52</td>
</tr>
<tr>
<td>CA_C0972</td>
<td>citC</td>
<td>Isocitrate dehydrogenase</td>
<td>NADH (104.6 ± 6.8), NADPH (7.1 ± 0.43)</td>
</tr>
<tr>
<td>CA_C1589</td>
<td>malS1</td>
<td>Malic enzyme</td>
<td>NADH (156 ± 9.6), NADPH (3.4 ± 0.24)</td>
</tr>
<tr>
<td>CA_C1596</td>
<td>malS2</td>
<td>Malic enzyme</td>
<td>NADH (142 ± 12.7), NADPH (2.9 ± 0.34)</td>
</tr>
</tbody>
</table>

*a One unit is the amount of enzyme that consumes 1 μmol of substrate per min.*
peptide labeling (36, 40) have been developed for two-dimensional high-performance liquid chromatography (2D methods using either 2D-protein gels (26) or peptide analysis by included from further analysis. The purpose of this study was also to quantify the number of cytoplasmic protein molecules per cell. Different quantitative methods using either 2D-protein gels (26) or peptide analysis by two-dimensional high-performance liquid chromatography (2D HPLC) coupled with tandem mass spectrometry (MS/MS) with peptide labeling (36, 40) have been developed for C. acetobutylicum. In collaboration with the Waters Company, we adapted a recently published method (64) using label-free peptide analysis after shotgun trypsin hydrolysis of cytosolic proteins. For approxi- mately 700 cytosolic proteins, it was possible to quantify the number of protein molecules per cell in at least one of the three steady states. This number is approximately 4 times higher than the number of cytosolic proteins detected in phosphate-limited acidogenic and solventogenic chemostat cultures by Janssen et al. (22) but similar to the number of cytosolic proteins detected by Venkataramanan et al. (40) by iTRAQ. Furthermore, the mini- mum number of protein molecules per cell detected was around 200 while the maximum number was approximately 300,000. For 96% of the cytosolic proteins that could be quantified, a linear relationship was obtained, with an \( R^2 \) value of >0.9, when the numbers of protein molecules per cell were plotted against the numbers of mRNA molecules per cell (see Data Set S2 in the sup- plemental material). This result indicated that for steady-state continuous cultures run at the same specific growth rate and with the same total amount of carbon supplied, the rate of protein turnover is proportional to the mRNA content for 96% of the genes. This result is not necessarily surprising, as it has previously been shown for other microorganisms such as E. coli (65) that the numbers of ribosomes and tRNAs per cell are dependent on the specific growth rate and not on the carbon source. The absolute protein synthesis rates for approximately 700 genes were calcu- lated by assuming that the rate of protein degradation is negligible compared to the rate of protein synthesis (see Data Set S2). These values varied from 0.0007 s\(^{-1}\) for CA_C3723 (sbb encoding a single-stranded DNA-binding protein) to 0.95 s\(^{-1}\) for CA_C0877 (cfa encoding a cyclopropane fatty acid synthase). Interestingly, the rate of protein synthesis appears to correlate inversely with the average number of mRNA molecules per cell (see Data Set S2).

(ii) Comparison of solventogenic versus acidogenic steady-state cells. Solventogenic cells were first comprehensively compared to acidogenic cells via quantitative transcriptomic and proteomic analyses. The complete transcriptomic and proteomic results are provided in Data Set S2 in the supplemental material. A similar study in phosphate-limited chemostat cultures was previously performed by Janssen et al. (22) using semiquantitative transcriptomic (two-color microarrays) and proteomic (2-DGE) methods. Among the 95 genes shown by Janssen et al. to be up-regulated, we qualitatively confirmed upregulation for 68; among the 53 genes shown by Janssen et al. to be down-regulated, we qualitatively confirmed downregulation for 27. What might explain the differences between the two studies? First, the culture conditions were slightly different in terms of dilution rate (0.075 h\(^{-1}\) for the work of Janssen et al., 0.05 h\(^{-1}\) in our study), phosphate limitation (0.5 mM for the work of Janssen et al., 0.7 mM in our study), and the pH of the acidogenic culture (5.7 for the work of Janssen et al., 6.3 in our study), leading to a larger amount of glucose consumed and thus a larger amount of products formed in our study. We are confident regarding the validity of our results because we found agreement quantitatively with the transcriptomic data whenever proteins were detected by our method, and thus, quantitative proteomic data were available. Below, we discuss these data in more detail, and striking differences in mRNA molecules per cell are highlighted in Fig. S2A in the supplemental material.

In total, 64 genes matched the significance criteria of \( \geq 4.0\)-fold-higher expression in solventogenesis than in acidogenesis as well as \( >0.2\) mRNA molecules per cell under at least one of the two conditions (see Table S2 in the supplemental material). In particular, high values (\(\sim 80\) to 150-fold) were documented for the sol operon genes (CA_P0162-CA_P0164) and confirmed by the proteomic analysis, in agreement with (i) the requirement of AdhE1 and CoA-transferase subunits for the production of solvents under solventogenic conditions (12, 47, 48, 66) and (ii) the previous study by Janssen et al. (22). Elevated upregulation (4- to 40-fold) of genes involved in serine biosynthesis (CA_C0014-CA_C0015), seryl-tRNA synthesis (CA_C0017), and arginine biosynthesis (CA_C2388) was detected at the mRNA level and confirmed by the proteomic analysis, in agreement with a previous metabolomic study in batch culture (67), which reported higher intracellular concentrations of serine and arginine in solventogenic cells. Interestingly, all these genes were previously shown to be upregulated in response to butanol stress (16), although these results were not confirmed by proteomic analysis (40). In addition, an \( \sim 4\)- to 8-fold upregulation of genes involved in purine biosynthesis (CA_C1392-CA_C1395, CA_C1655, and CA_C2445) was detected at the mRNA level and confirmed by the proteomic analysis. Similar to the results in the study by Janssen et al. (22), an \( \sim 5\)-fold upregulation of a gluconate dehydrogenase (CA_C2607) was detected; however, as this protein was not detected, this was not confirmed by proteomic analysis.

As reported in previous studies (22, 37), we found elevated upregulations (\( \sim 4\)- to 16-fold) of the genes involved in the production of (i) a nonfunctional cellulosome (CA_C0910- CA_C0918 and CA_C0561) (20, 68) and (ii) noncellulosomal peptate lyase-encoding genes (CA_P0056 and CA_C0574) at the mRNA level. However, these results could not be verified by proteomic analysis, as exoproteome analysis was not performed in

Yoo et al.
this study. All these genes, except CA_P0056, were also shown to be upregulated in response to a butanol stress (16).

Importantly, spo0A (CA_C2071), encoding a regulator of sporulation and solvent production (69–71), showed an increase in expression at the level of both mRNA and protein molecules per cell. This increased expression does not agree with previous chemostat culture studies by Grimmelmer et al. (37) and Janssen et al. (22) but does agree with batch culture studies (21, 25) and also supports the common notion of Spo0A acting as a master regulator of solventogenesis. hsp18 (CA_C3714), encoding a gene product involved in solvent tolerance (72), also exhibited an ~4.5-fold increase in mRNA and protein molecules per cell, in agreement with a previous butanol stress study (40). A striking difference between the study by Janssen et al. and ours was observed with regard to the level of this chaperone, which in contrast to our study showing an ~4.5-fold increase under solventogenesis, was decreased (~5-fold) in the study by Janssen et al. (22). Nonetheless, this difference appears to be due to the limitation of 2-DGE, because 3 different proteins could be detected in the "Hsp18 spot" and transcriptional changes in hsp18 did not correlate with the proteomic data (22); in contrast, our quantitative transcriptomic and proteomic data showed good correlation ($R^2 > 0.9$).

The detailed results of the 45 ORFs that exhibited a ~4.0-fold decreases in numbers of mRNA molecules per cell under solventogenic versus aciogenic conditions and of a number with >0.2 mRNA molecules per cell under at least one of the two conditions are given in Table S2 in the supplemental material. Significantly, in this metabolic state, various genes involved in the assimilation of different carbon sources were downregulated. For example, the highest decrease (~6–70-fold) at the mRNA level was observed for genes (CA_C0422–CA_C0426) involved in sucrose transport, metabolism, and the regulation of these genes, which was confirmed by the proteomic analysis. In addition, two genes involved in mannans (CA_C0332) and maltose (CA_C0533) metabolism exhibited 4- and 10-fold decreases, respectively, in their mRNA levels. Because acidogenic culture reached glucose limitation but a small amount of glucose remained in solventogenic culture (similar to our previous publication [9]), this phenomenon can be explained by a release of catabolite repression in acidogenic cultures. The similar high expression observed for CA_C0422–CA_C0426, CA_C0332, and CA_C0533 in acidogenic and acidogenic cultures that were glucose limited is in agreement with this hypothesis. Two genes located on the megaplasmid pSOL1 (CA_P0036 and CA_P0037), encoding a cytosolic protein of unknown function and a potential transcriptional regulator, respectively, exhibited particularly high scores corresponding to an ~60- to 70-fold decrease, which is in good agreement with the proteomic data and the previous study by Janssen et al. (22). Interestingly, under all conditions, these two proteins are present at a 1:1 molar ratio. Furthermore, three genes involved in cysteine (CA_C2783) and methionine (CA_C1825 and CA_C0390) biosynthesis exhibited ~5-fold decreases in their numbers of mRNA and protein molecules per cell in agreement with a previous metabolomics study by Amador-Noguez et al. (67), showing an ~5-fold decrease in intracellular methionine in solventogenesis.

(iii) Comparison of alcohologenic versus aciogenic steady-state cells. Alcohologenic cells were comprehensively compared to acidogenic cells by quantitative transcriptomic and proteomic analyses. The complete transcriptomic results are listed in Data Set S2, and striking differences are highlighted in Fig. S2B, both in the supplemental material. In total, 52 genes matched the significance criterion of $\geq$4.0-fold-higher expression in alcohologenesis than in acidogenesis as well as $\geq$0.2 mRNA molecules per cell under at least one of the two conditions (see Table S3). In particular, high values (~55- to 520-fold) were documented for the gene cluster coding for glycerol transport and utilization (CA_C1319-CA_C1323) and confirmed by the proteomic analysis, in agreement with the requirement of GlpK (glycerol kinase) and GlpAB (glycerol-3-phosphate dehydrogenase) for glycerol utilization in alcohologenetic metabolism (6, 9). High upregulation (160-fold) of adhE2 (CA_P0035), which is involved in alcohol production under alcohologenic conditions (66), was detected and correlated with a high AdhE2 protein concentration. Interestingly, CA_C3486, which encodes a multimeric flavodoxin, was also highly expressed (~6-fold) and may participate in redistribution of the electron flux in favor of butanol under alcohologenic conditions. Of note, ~20- to 70-fold upregulation of a gene cluster involved in sulfate transport, reduction, and incorporation to produce cysteine (CA_C0102-CA_C0110); ~4-fold upregulation of cysK (CA_C2235), which is also involved in cysteine synthesis; and ~7- to 10-fold upregulation of an operon (CA_C3325-CA_C3327) involved in cysteine transport were detected at the mRNA level and confirmed by the proteomic analysis for the cystolic proteins detected (CA_C0102-CA_C0104, CA_C0107, CA_C0109-CA_C0110, CA_C2235, and CA_C3327). All of these genes/operons were shown to possess a CymR-binding site in their promoter regions, and some have been shown to be upregulated in response to butanol stress (16).

An ~3- to 5-fold upregulation of an operon involved in histidine synthesis and histidyl-tRNA synthesis (CA_C0935-CA_C0943) and 5-fold upregulation of a gene involved in arginine biosynthesis (CA_C2388) were also detected at the mRNA level and confirmed by the proteomic analysis. These genes were also shown to be upregulated under solventogenic conditions (this study) and in response to butanol stress (16).

The detailed results of the 64 ORFs that exhibited a $\geq$4.0-fold decrease in transcript levels under aciogenic versus acidogenic conditions and $\geq$0.2 mRNA molecules per cell under at least one of the two conditions are given in Table S3 in the supplemental material. The highest decrease (~70-fold) at the mRNA level was observed for an operon (CA_C0427-CA_C0430) involved in glycerol-3-phosphate transport and coding for a glycerophosphoryl diester phosphodiesterase, which was confirmed by the cytosolic protein analysis. As observed under solventogenic conditions, CA_P0036 and CA_P0037 exhibited ~40- to 50-fold-lower expression levels, which agrees well with the proteomic data. Furthermore, an operon involved in phosphate uptake (CA_C1705-CA_C1709), an operon encoding an indolepyruvate ferredoxin oxidoreductase (CA_C2000-CA_C2001), and a gene encoding a pyruvate decarboxylase (CA_P0025) exhibited ~80- to 350-fold, ~4- to 5-fold, and ~4-fold decreases, respectively, at the mRNA level, confirmed by the proteomic analysis. Additionally, two clusters of genes involved in fatty acid biosynthesis/ degradation (CA_C2004-CA_C2017) exhibited ~3.5- to 6-fold decreases at the mRNA level, a result that could not be confirmed by the proteomic analysis as the corresponding proteins were not detected.

Metabolic flux analysis of C. acetobutylicum under stable acidogenic, solventogenic, and alcohologenic conditions. To perform a metabolic flux analysis of C. acetobutylicum under stable acidogenic, solventogenic, and alcohologenic conditions,
Cac967 was combined with our transcriptomic and proteomic data. As a first simple example, we present how the gene responsible for pyruvate ferredoxin oxidoreductase (PFOR) activity was identified. This gene encodes a key enzyme in the glycolytic pathway that decarboxylates pyruvate to produce reduced ferredoxin, CO₂, and acetyl-CoA. Two putative PFOR-encoding genes (CA_C2229 and CA_C2499) were identified in our GSM (see Data Set S1 in the supplemental material). Under all conditions, only CA_C2229 was transcribed (average of 56 mRNA molecules per cell) and translated (average of 166,000 protein molecules per cell).

As a second simple example, we present how the main enzyme responsible for crotonyl-CoA reduction to butyryl-CoA was identified. Two different enzymes can potentially catalyze this reaction: the BCD complex encoded by bcd, etfB, and etfA (CA_C2711, CA_C2710, and CA_C2709, respectively) which consumes 2 moles of NADH and produces 1 mole of reduced ferredoxin (see Fig. S1 in the supplemental material) and TER (trans-2-enoyl-CoA reductase) encoded by CA_C0642, which consumes only 1 mole of NADH (73). Under all conditions, bcd was much more highly transcribed than CA_C0642 (67 versus 1.2 mRNA molecules per cell) and translated (average of 166,000 protein molecules per cell). Under all conditions, bcd was much more highly transcribed than CA_C0642 (67 versus 1.2 mRNA molecules per cell) and translated (average of 166,000 protein molecules per cell).

As a complex example, we also present the actors in the different butanol pathways and their cofactor specificities. Five proteins could potentially be involved in the last two steps of butanol formation. AdhE1 retains only NADH-dependent aldehyde dehydrogenase activity, whereas AdhE2 is a bifunctional NADH-dependent aldehyde-alcohol dehydrogenase (66); BdhA, BdhB, and BdhC are NADPH-dependent alcohol dehydrogenases. For each of the three conditions and for each of the abovementioned genes and their corresponding proteins, the number of mRNA molecules per cell and the number of protein molecules per cell were measured. The percentage of the total butanol flux due to each of the five enzymes was calculated by assuming that all five enzymes function at their V_max and using the amount of each protein per cell. The results are presented in Fig. 1. Under acidogetic conditions, the entire butyraldehyde dehydrogenase flux is due to AdhE2, whereas the butanol dehydrogenase flux is primarily due to BdhB and BdhA. Under solventogenic conditions, the butyraldehyde dehydrogenase flux is largely due to AdhE1, whereas the butanol dehydrogenase flux is primarily due to BdhB, BdhA, and BdhC, in decreasing order of activity. Finally, under alcohologenic conditions, all of the flux of butyraldehyde dehydrogenase activity and most of that of butanol dehydrogenase activity are due to AdhE2. In summary, the last two steps of butanol production consume 1 mole of NADH and 1 mole of NADPH.

![Butanol pathway analysis under acidogenesis (A), solventogenesis (B), and alcohologenesis (C). (Left) Numbers of mRNA (blue) and protein (green) molecules per cell for the five enzymes potentially involved in butanol production. (Right) Activity distributions of the five enzymes are shown for each step under the arrows. The primary cofactors used for each step are shown over the arrows. Butanol flux is indicated under the word “Butanol.”](mbio.asm.org/FIG1.png)
under acidogenic and solventogenic conditions and 2 moles of NADH under alcohologenic conditions (Fig. 1). These results have strong implications for the distribution of electron fluxes and the use of reduced ferredoxin under the respective studied conditions. Under acidogenic conditions, reduced ferredoxin is primarily used to produce hydrogen, and only a small fraction is used to produce the NADH needed for butyrate formation and the NADPH needed for anabolic reactions (Fig. 2A). However, under alcohologenic conditions, reduced ferredoxin is primarily used to produce the NADH needed for alcohol formation (Fig. 2C); under solventogenic conditions, although reduced ferredoxin is predominantly utilized for hydrogen production, a significant amount is used for the NADPH formation needed for the final step of alcohol formation by BdhB, BdhA, and BdhC, as C. acetobutylicum has no oxidative pentose phosphate pathway (zwf, encoding glucose 6-phosphate-dehydrogenase, is absent) to produce NADPH (Fig. 2B and 3). Although the enzymes converting reduced ferredoxin to NADPH or NADH, namely, ferredoxin-NAD+ reductase and ferredoxin-NAD+ reductase, and their corresponding genes are unknown, they likely play key roles in alcohol formation under solventogenic and alcohologenic conditions, respectively.

A fourth example of metabolic flux analysis is the identification of the hydrogen production pathway. Three hydrogenases are potentially involved: two Fe–Fe hydrogenases, HydA (encoded by CA_C0028) and HydB (encoded by CA_C3230), and one Ni–Fe hydrogenase, HupSL (encoded by CA_P0141-CA_P0142). The hydB and the hupSL genes are not expressed under all three conditions, nor were the HydB and HupSL proteins detected by quantitative proteomic analysis. As HydA is the only hydrogenase present, how can the lower observed flux in H2 production under solventogenic and alcohologenic conditions (compared to acidogenic conditions) be explained? Under solventogenic conditions, there is a 3-fold decrease in the expression of hydA; this is associated with a 2-fold decrease in the expression of fdx1 (CA_C0303), which encodes the primary ferredoxin, the key redox partner for the hydrogenase. As these results were confirmed by the proteomic analysis, they may explain the 1.3-fold decrease in H2 production under solventogenic conditions compared to alcohologenic conditions (Fig. 2B). Nonetheless, under alcohologenic conditions a 1.7-fold decrease in H2 production (compared to acidogenic conditions) is associated with a 1.8-fold-higher expression of hydA, a 3-fold decrease in the expression of fdx1, and a 6-fold increase in the expression of CA_C3486, which encodes a multimeric flavodoxin, another potential redox partner for the hydrogenase. In fact, the reduced multimeric flavodoxin may be a better substrate for the ferredoxin-NAD+ reductase than for the primary hydrogenase, as was previously shown for reduced neutral red (8). This result would explain the low flux in hydrogen production and the high flux in ferredoxin-NAD+ reductase production under alcohologenic metabolism obtained through growth either in glucose-glycerol mixtures or in glucose in the presence of neutral red (8).

A fifth example of metabolic flux analysis is the glyceraldehyde-3-phosphate oxidation pathway. Two glyceraldehyde-3-phosphate dehydrogenases are potentially involved: GapC (encoded by CA_C0709) (74), which phosphorylates and produces NADH, and GapN (encoded by CA_C3657) (75), which is nonphosphorylating and produces NADPH. For each of the three conditions and each of the genes studied, the numbers of mRNA molecules and protein molecules per cell were measured. The percentage of the total glycolytic flux due to each of the enzymes was calculated by assuming that both enzymes function at their previously published Vmax levels (74, 75) and using the amount of each protein per cell. Here, results are presented for only solventogenic metabolism, though qualitatively, the conclusions were the same for all conditions: gapN is poorly expressed compared to gapC (0.56 versus 66 mRNA molecules per cell; 3,500 versus 190,000 protein molecules per cell) (see Data Set S2 in the supplemental material), and GapN would be responsible for less than 5% of the glycolytic flux.

Two fluxes involved in anaplerotic reactions, namely, those for pyruvate carboxylase (encoded by CA_C2660) and NADH-dependent malic enzymes (encoded by CA_C1589 and CA_C1596), could not be solved using our GSM analysis coupled with transcriptomic and proteomic analyses. All of the genes studied were transcribed and translated under all conditions, and because all fermentations occurred under a high partial pressure of CO2, malic enzymes could function in both malate production and the high flux in ferredoxin-NAD+ reductase production under alcohologenic metabolism obtained through growth either in glucose-glycerol mixtures or in glucose in the presence of neutral red (8).

A fourth example of metabolic flux analysis is the identification of the hydrogen production pathway. Three hydrogenases are potentially involved: two Fe–Fe hydrogenases, HydA (encoded by CA_C0028) and HydB (encoded by CA_C3230), and one Ni–Fe hydrogenase, HupSL (encoded by CA_P0141-CA_P0142). The hydB and the hupSL genes are not expressed under all three conditions, nor were the HydB and HupSL proteins detected by quantitative proteomic analysis. As HydA is the only hydrogenase present, how can the lower observed flux in H2 production under solventogenic and alcohologenic conditions (compared to acidogenic conditions) be explained? Under solventogenic conditions, there is a 3-fold decrease in the expression of hydA; this is associated with a 2-fold decrease in the expression of fdx1 (CA_C0303), which encodes the primary ferredoxin, the key redox partner for the hydrogenase. As these results were confirmed by the proteomic analysis, they may explain the 1.3-fold decrease in H2 production under solventogenic conditions compared to alcohologenic conditions (Fig. 2B). Nonetheless, under alcohologenic conditions a 1.7-fold decrease in H2 production (compared to acidogenic conditions) is associated with a 1.8-fold-higher expression of hydA, a 3-fold decrease in the expression of fdx1, and a 6-fold increase in the expression of CA_C3486, which encodes a multimeric flavodoxin, another potential redox partner for the hydrogenase. In fact, the reduced multimeric flavodoxin may be a better substrate for the ferredoxin-NAD+ reductase than for the primary hydrogenase, as was previously shown for reduced neutral red (8). This result would explain the low flux in hydrogen production and the high flux in ferredoxin-NAD+ reductase production under alcohologenic metabolism obtained through growth either in glucose-glycerol mixtures or in glucose in the presence of neutral red (8).
material, the anaplerotic fluxes presented are net anaplerotic fluxes, which were attributed to pyruvate carboxylase.

The flux in the oxidative branch of the TCA cycle was much higher than that in the reductive branch (Fig. 3; see also Fig. S3 in the supplemental material). In agreement with the $^{13}$C labeling flux data reported by Amador-Noguez et al. (55), who demonstrated the flux from oxaloacetate to malate, but in contrast to the report by Au et al. (76), in which no flux could be measured through this enzyme, under all three conditions, we measured ~1,000 malate dehydrogenase (CA_C0566) protein molecules per cell that could catalyze the first step of the TCA reductive branch (see Data Set S2).

**Conclusion.** In this work, an improved GSM containing new and validated biochemical data was developed in conjunction with quantitative transcriptomic and proteomic analyses to obtain accurate fluxomic data. These “omics” data allowed for (i) the determination of the distribution of carbon and electron fluxes, (ii) the elucidation of the different genes/enzymes involved in the primary metabolism of *C. acetobutylicum*, and (iii) a better understanding of the regulation of *C. acetobutylicum* primary metabolism under different physiological conditions. The information provided in this study will be important for the further metabolic engineering of *C. acetobutylicum* to develop a commercial process for the production of n-butanol.

**MATERIALS AND METHODS**

**Chemicals and other reagents.** All chemicals were of reagent grade and were purchased from Sigma-Aldrich Chimie (Saint-Quentin Fallavier, France) or from VWR Prolabo (Fontenay sous Bois, France). All gases used for gas flushing of the medium and for the anaerobic chamber were of the highest purity available and were obtained from Air Liquide (Paris, France). All restriction enzymes and Crimson Taq DNA polymerase used for colony PCR were supplied by New England Biolabs (MA, USA) and were used according to the manufacturer’s instructions. DNA fragments for vector constructions were amplified using Phusion high-fidelity DNA polymerase (New England Biolabs).

**Culture conditions. (i) Batch culture.** All liquid cultures of *C. acetobutylicum* ATCC 824 ΔCA_C1502 Δupp (P. Soucaille, R. Figge, and R. Croux, 2014, U.S. Patent 8,628,967) were performed in 30-ml or 60-ml glass vials under strict anaerobic conditions in clostridium growth medium (CGM) as described previously (77) or in synthetic medium (MS) as...
described previously (6). *C. acetobutylicum* was stored in spore form at −20°C after sporulation in MS medium. Heat shock was performed for spore germination by immersing the bottle in a water bath at 80°C for 15 min.

(ii) Continuous culture. The conditions described previously by Vasconcelos et al. (6) and Girbal et al. (9) were used for the phosphate-limited continuous culture of *C. acetobutylicum* fed a constant total carbon amount of 995 mM. The cultures were maintained under acidogenesis (pH 6.3, 995 mM carbon from glucose), solvenogenes (pH 4.4, 995 mM carbon from glucose), and alcogenogenesis (pH 6.3, 498 mM carbon from glucose and 498 mM carbon from glycerol).

**RNA extraction and microarray.** For transcriptomic analysis, 3-ml samples were collected from chemostat cultures and immediately frozen in liquid nitrogen. The frozen cell cultures were ground promptly with 2-mercaptoethanol in a liquid nitrogen-cooled mortar. RNA was extracted using an RNeasy midikit (Qiagen, Courtabeuf, France) according to the manufacturer’s instructions with the supplementation of DNase treatment using the RNase-free DNase set (Qiagen). RNA quantity and composition were analyzed using an Agilent 2100 bioanalyzer (Agilent Technologies, Massy, France) and a NanoDrop ND-1000 spectrophotometer (Labtech France, Paris, France) at 260 nm and 280 nm. All microarray procedures were performed according to the manufacturer’s protocol (Agilent one-color microarray-based exon analysis). Briefly, the RNAs were labeled with a low-input Quick Amp labeling kit and hybridized following a one-color microarray-based gene expression analysis protocol. The slides were scanned using a Tecan MS200 scanner and analyzed using Feature Extraction V.11.5.1.1.

**Protein extraction and analysis.** For proteomic analysis, 20-ml samples were collected from chemostat cultures and treated according to the standard operating procedures developed by Schwarz et al. (78) for the extraction of intracellular proteins, except that phenylmethylsulfonyl fluoride (PMSE) was not added. Samples of 200 µg of each of the lyophilized protein extracts were dissolved at 80°C in 100 µl of 0.1% RapiGest (Waters) in water. Disulfide bonds were reduced with the addition of dithiothreitol (DTT) at 2 mM and incubation at 60°C for 15 min. Cysteine residues were carboxyamidomethylated with the addition of iodoacetamide to a concentration of 10 mM and incubated in the dark at room temperature. Trypsyl digestion was performed with trypsin (10 µg/ml) at 37°C for 12 h. Protein hydrolysates were acidified with 5 f مل/ل for 2 min to remove the RapiGest precipitate. The supernatant was collected. Postdigestion samples at a concentration of 10 mg/ml were mixed at a ratio of 1:1 with 40 fمل/ل (5% formic acid) to Minyeong Yoo.

**Biomass concentration**

Biomass concentration was determined both by counting the number of cells per milliliter, as previously described (81), and by the DCW method after centrifugation (16,000×g, 5 min, room temperature), two washes with Milli-Q water, and drying under vacuum at 80°C. The concentrations of glucose, glycogen, acetate, butyrate, lactate, pyruvate, acetoin, acetone, ethanol, and butanol were determined based on high-performance liquid chromatography (HPLC), as described by Dusseaux et al. (82), except that the concentration of HSO₄ was changed to 0.5 mM, as required for mobile phase optimization. The concentrations of formate and fumarate were measured using a formate assay kit (Sigma-Aldrich) and a fumarate assay kit (Sigma-Aldrich), according to the manufacturer’s instructions.

**Measurement of fermentation parameters.** Biomass concentration was determined by counting the number of cells per milliliter, as previously described (81), and by the DCW method after centrifugation (16,000×g, 5 min, room temperature), two washes with Milli-Q water, and drying under vacuum at 80°C. The concentrations of glucose, glycogen, acetate, butyrate, lactate, pyruvate, acetoin, acetone, ethanol, and butanol were determined based on high-performance liquid chromatography (HPLC), as described by Dusseaux et al. (82), except that the concentration of HSO₄ was changed to 0.5 mM, as required for mobile phase optimization. The concentrations of formate and fumarate were measured using a formate assay kit (Sigma-Aldrich) and a fumarate assay kit (Sigma-Aldrich), according to the manufacturer’s instructions.

**Metabolic enzyme expression and purification.** Information on metabolite expression and purification is provided as Text S1 in the supplemental material.

**Microarray data accession number.** The microarray data can be accessed at GEO through accession number GSE69973

---

**SUPPLEMENTAL MATERIAL**


Text S1, DOCX file, 0.03 MB.
Data Set S1, XLSX file, 0.2 MB.
Data Set S2, XLSX file, 2.1 MB.
Figure S1, PDF file, 0.03 MB.
Figure S2, PDF file, 0.2 MB.
Figure S3, PDF file, 0.1 MB.
Figure S4, PDF file, 0.1 MB.
Table S1, DOCX file, 0.01 MB.
Table S2, DOCX file, 0.03 MB.
Table S3, DOCX file, 0.03 MB.

**ACKNOWLEDGMENTS**

We thank Jean-Louis Uribelarrea, Sophie Lamarre, and Lidwine Trouilh for help with the data analysis.

This work was financially supported by the European Community’s Seventh Framework Program “CLOSTNET” (PEOPLE-ITN-2008-237942) to Minyeong Yoo.

**REFERENCES**


