Research Article

Genome analysis of a hyper acetone-butanol-ethanol (ABE) producing *Clostridium acetobutylicum* BKM19

Changhee Cho[1], Donghui Choe[2], Yu-Sin Jang[3], Kyung-Jin Kim[4], Won Jun Kim[1], Byung-Kwan Cho[2,5], E. Terry Papoutsakis[6], George N. Bennett[7], Do Young Seung[8], and Sang Yup Lee[1,*]

[1]Department of Chemical and Biomolecular Engineering (BK21 Plus Program), BioProcess Engineering Research Center, Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

[2]Department of Biological Sciences and KI for the BioCentury, KAIST, Daejeon, Republic of Korea

[3]Department of Agricultural Chemistry and Food Science, Gyeongsang National University, Jinju, Gyeongsangnam-do, Republic of Korea

[4]School of Life Sciences, KNU Creative BioResearch Group, Kyungpook National University, Daegu, Korea

[5]Intelligent Synthetic Biology Center, KAIST, Daejeon, Republic of Korea

[6]Department of Chemical Engineering and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware, USA

Submitted: 31-Jul-2016
Revised: 10-Nov-2016
Accepted: 02-Dec-2016

[7]Department of Biochemistry and Cell Biology, Rice University, Houston, Texas, USA

[8]GS Caltex Corporation R&D Center, Daejeon, Republic of Korea

Correspondence: Prof. Sang Yup Lee, Department of Chemical and Biomolecular Engineering, KAIST, Daejeon 34141, Republic of Korea

E-mail: leesy@kaist.ac.kr

Keywords: *Clostridium acetobutylicum* · Butanol · Thiolase

Abbreviations: **ABE,** acetone-butanol-ethanol; **SNVs,** single nucleotide variants; **InDels,** insertions and deletions; **CGM,** clostridial growth medium; **RDR,** regulatory determinant region

Abstract

We previously reported development of a hyper acetone-butanol-ethanol (ABE) producing *Clostridium acetobutylicum* BKM19 strain capable of producing 30.5% more total solvent by random mutagenesis of its parental strain PJC4BK, which is a *buk* mutant *C. acetobutylicum* ATCC 824 strain. Here BKM19 and PJC4BK strains were re-sequenced by a high-throughput sequencing technique to understand the mutations responsible for enhanced solvent production. In comparison with the *C. acetobutylicum* PJC4BK, 13 single nucleotide variants (SNVs), 1 deletion and 1 back mutation SNV were identified in the *C. acetobutylicum* BKM19 genome. Except for 1 SNV found in the megaplasmid, all mutations were found in the chromosome of BKM19. Among them, a mutation in the *thlA* gene encoding thiolase was further studied with respect to enzyme activity and butanol production. The mutant thiolase (*thlA*^{V5A}) showed a 32% higher activity than that of the wild-type thiolase (*thlA*^{WT}). In batch fermentation, butanol production was increased by 26% and 23% when the *thlA*^{V5A} gene was overexpressed in the wild-type *C. acetobutylicum* ATCC 824 and in its derivative, the *thlA*-knockdown TKW-A strain, respectively. Based on structural analysis, the mutation in thiolase does not have a direct effect on the regulatory determinant region (RDR). However, the mutation at the 5th residue seems to influence the stability of the RDR, and thus, increases the enzymatic activity and enhances solvent production in the BKM19 strain.

1 Introduction

There has been much interest in producing biofuels from renewable non-food biomass to cope with increasing pressure on our environment and climate change [1-3]. Butanol, already being an important industrial solvent in various applications, has been considered as one of the better biofuels than bioethanol due to its similar fuel properties to gasoline and less hygroscopy [4-6]. Although butanol is currently produced from fossil fuels by a chemical process, it can also be produced through acetone-butanol-ethanol (ABE) fermentation using solventogenic clostridia [7, 8].

In the industrial ABE process, *Clostridium acetobutylicum* has most widely been employed [9-11]. Recently, several high performance *C. acetobutylicum* strains have been developed for enhanced production of butanol through metabolic engineering [12-16]. At the same time, random mutation and screening methods have also been used to develop a hyper ABE producer [16-18]. In order to understand the improved phenotypic changes in these randomly mutated strains, a high-throughput genome sequencing technique can be employed [19].

We have recently reported the development of the *C. acetobutylicum* BKM19 strain by random mutagenesis of PJC4BK strain, which is a *buk* mutant *C. acetobutylicum* ATCC 824 strain [17]. The BKM19 strain is capable of producing 32.5 g/L of ABE (17.6, 10.5, 4.4 g/L of butanol, ethanol, and acetone, respectively) from 85.2 g/L of glucose in batch fermentation [17]. In this study, this hyper ABE producing BKM19 and its parental strain PJC4BK were sequenced with a high-throughput sequencing technique, and were compared for identifying the mutations responsible for enhanced solvent production. Multiple single nucleotide variants (SNVs) together with deletion and back mutation were found and analyzed. Among several SNVs, we focused one key mutation in the *thlA* gene, which encodes thiolase. Thiolase is the first enzyme of the butyrate/butanol fermentative pathway, catalyzing the condensation of two acetyl-CoA (C2) to acetoacetyl-CoA (C4). As a matter of priority, thiolase was further analyzed through construction of engineered strains and cultivation studies.

2 Materials and methods

2.1 Bacterial strains and extraction of genomic DNA

A single colony of *Clostridium acetobutylicum* strain (PJC4BK [20] and BKM19 [17]) was inoculated in clostridial growth medium (CGM) containing (per L) 0.75 g KH_2PO_4, 0.75 g K_2HPO_4, 1.0 g NaCl, 0.017 g $MnSO_4 \cdot 5H_2O$, 0.70 g $MgSO_4 \cdot 7H_2O$, 0.01 g $FeSO_4 \cdot 7H_2O$, 2.0 g L-asparagine, 5.0 g yeast extract, 2.0 g $(NH_4)_2SO_4$, and 80 g glucose. When the cell density reached an OD_{600} of 1.0, genomic DNA was extracted with Exgene^{TM} cell SV (GeneAll, Seoul, Korea).

2.2 Next generation sequencing analysis

The genomic DNAs of the PJC4BK and BKM19 strains were fragmented by the Covaris S220 system (Thermo Fisher Scientific, MA, USA). The sequencing library was constructed from the fragmented genomic DNA with the TruSeq DNA LT Sample Prep Kit (Illumina, CA, UST) following the manufacturer's instruction. The constructed libraries were sequenced on an Illumina MiSeq V2 with a read length of 51 bp. A total of 12,571,501 and 12,537,829 reads were generated for the PJC4BK and BKM19 genomes, respectively.

Sequencing results were analyzed with the CLC Genomics Workbench (CLC Bio) following the analysis pipeline. First, raw reads were quality trimmed with a quality limit score of 0.05 and reads with more than 2 ambiguous nucleotides were discarded. Quality trimmed reads were mapped onto the chromosomal DNA sequence of *C. acetobutylicum* (NC_003030.1) with the following mapping parameters: similarity fraction of 0.9, length fraction of 0.9, mismatch cost of 2, insertion and deletion cost of 3, and random mapping of non-specific matches. Those reads that were not mapped onto the chromosomal DNA were mapped onto the megaplasmid (NC_001988.2) using the same parameters as above. A total of 589,403,648 and 587,463,839 bases were mapped for the PJC4BK and BKM19 strains, respectively, providing genome coverage of approximately 150-fold. For the megaplasmid, 36,689,038 and 38,452,460 bases were generated from PJC4BK and BKM19,

respectively, corresponding to a coverage of over 190-fold.

Sequence variants were detected by the Quality-based Variant Detection toolbox of the CLC Genomics Workbench, using the following parameters with the bacterial genetic code: neighborhood radius of 5, maximum gap and mismatch count of 5, minimum neighborhood quality of 30, minimum central quality of 30, minimum coverage of 10, minimum variant frequency of 10 %, and maximum expected alleles of 4. The non-specific matches were ignored. Among the detected variants, tRNA, and rRNA, variants with low evidence (low coverage region), and error-prone regions (consecutive bases, such as TTTTTTTT) were discarded during the manual curation process.

2.3 Sanger sequencing for the confirmation of SNVs/InDels

About 500-base genomic regions around the SNVs and insertions and deletions (InDels) identified by the mapping program were amplified by PCR and sequenced on a capillary sequencer with the Sanger method using the commercial sequence services of Macrogen (Seoul, Korea).

2.4 Plasmid construction and transformation

Plasmids and primers used in this study are listed in Table I. All restriction enzymes were purchased from New England BioLabs (Ipswich, MA, USA). T4 DNA ligase (Elpisbio, Daejeon, Korea) was used for ligation, and a proofreading *Pfu-x* DNA polymerase (Solgent, Daejeon, Korea) was used for PCR. Plasmid pTHL1-Cm [21] was used as a parent vector for cloning. The *thlA* and mutant $thlA^{V5A}$ genes were amplified from the genomic DNA of *C. acetobutylicum* ATCC 824 with primer 1-primer 3 and primer 2-primer 3, respectively. The PCR products were cloned into the *Pst*I and *Ava*I sites of plasmid pTHL1-Cm to construct pTHL1-ThlA and pTHL1-V5A, respectively. To overcome the clostridial restriction system [22], the plasmids were methylated prior to transformation into *C. acetobutylicum* ATCC 824 and the *thlA*-knockdown mutant TKW-A strain.

The methylated plasmids were transformed into *C. acetobutylicum* ATCC 824 and TKW-A strain by electroporation (Bio-Rad, Hercules, CA, USA; 2.5 kV, ∞ Ω, and 25 μF).

Plasmid pET-30a(+) was used to express protein in *Escherichia coli*. Using primer pairs primer 4-primer 6 and primer 5-primer 6, the *thlA* and mutant $thlA^{V5A}$ genes, respectively, were amplified from *C. acetobutylicum* genomic DNA. The amplified products were digested with *Nco*I and *Ava*I and ligated with pET-30a (+) also digested with *Nco*I and *Ava*I to construct pET-ThlA and pET-V5A, respectively. The constructed vectors were transformed into *E. coli* BL21(DE3) by electroporation (Bio-Rad, Hercules, CA, USA; 1.8 kV, 200 Ω, and 25 μF).

2.5 Expression and purification of wild-type and mutant thiolases

Plasmids pET-ThlA and pET-V5A were introduced into *E. coli* BL21(DE3), and the transformants were cultured in 10 mL of LB broth containing 100 μg/mL kanamycin at 37ºC. Then, 1 mL of culture was inoculated into 100 mL of LB broth containing 100 μg/mL kanamycin and cultured at 37ºC until reaching an OD_{600} of 0.6 - 0.7. Then, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to each flask to a final concentration of 1 mM, and the flasks were incubated for another 4 h. Cells were collected by centrifugation and resuspended in 50 mL of binding buffer (50mM sodium phosphate, 300mM NaCl, pH 7.0), and were disrupted by ultrasonication (VCX 600, Sonics & Materials Inc., Newtown, CT, USA). The cell extract was centrifuged at 10,000 x g for 10 min at 4ºC. The supernatant was filtered using a 0.45 μm pore size filter (Satorius, Göttingen, Germany) and subjected to TALON® metal affinity purification (Clontech Laboratories, Inc., CA, USA)

2.6 Enzyme assay

After purification of the wild-type thiolase ($ThlA^{WT}$) and mutant thiolase ($ThlA^{V5A}$), the enzyme activities were measured by monitoring the decrease of acetoacetyl-CoA absorbance at 303 nm. The reaction mixture (1 mL) contained 100 mM Tris-HCl (pH 8.0), 1 mM 1,4-dithiothreitol, 10mM

$MgCl_2$, 0.05 mM CoA and 0.05 mM acetoacetyl-CoA, to which 0.2 ug of wild-type or mutant thiolase protein was added to start the reaction. The molar extinction coefficient of 14,000 M^{-1} cm^{-1} was used [23].

2.7 Batch fermentation

Batch fermentation was performed in a 5.0 L LiFlus GX bioreactor (Biotron, Gyeonggi-Do, Korea) containing 1.8 L of CGM supplemented with 80 g/L glucose. A single colony was inoculated into a test tube containing 10 mL of CGM and grown overnight. When the cell density reached an OD_{600} of 1.0 – 2.0, it was transferred into a 500 mL flask containing 200 mL of CGM supplemented with 80 g/L of glucose. At the OD_{600} of 2.0, cells were inoculated into the bioreactor. The culture pH was adjusted above 5.0 with 14% (v/v) ammonia solution. The bioreactor was operated at 37°C and 200 rpm, and oxygen-free nitrogen gas was sparged at a flow rate of 0.5 L/min after passing through an oxygen trap (Agilent, Santa Clara, CA, USA).

2.8 Analytical techniques

Cell growth was monitored by measuring the optical density at 600 nm using an Ultraspec 3100 Pro spectrophotometer (Amersham Biosciences, Uppsala, Sweden). The concentrations of solvents including acetone, ethanol and butanol were analyzed by gas chromatography (7890; Agilent Technologies, CA, USA) equipped with an 80/120 Carbopack B AW packed glass column (Supelco, PA, USA) and a flame ionization detector (FID; Agilent Technologies). The concentrations of residual glucose and organic acids, including acetic acid, butyric acid, and lactic acid, were analyzed by high performance liquid chromatography (Waters 1515/2414/2707, Waters, Milford, MA, USA). A MetaCarb 87H column (300 by 7.8 mm; Agilent) was eluted isocratically with 0.01 N H_2SO_4 at 35 °C at a flow rate of 0.5 mL/min.

2.9 *In silico* metabolic flux analysis

The *in silico* metabolic flux analysis was conducted to predict metabolic fluxes of *C. acetobutylicum* ATCC 824 strain overexpressing the $thlA^{WT}$ and $thlA^{V5A}$ genes using the genome-scale metabolic model of *C. acetobutylicum* ATCC 824, iCac802 [24]. The glucose uptake rates and byproduct formation rates were used as constraints for the model, which is experimentally measured values at the exponential phase. Flux values (i.e., minimum and maximum values) for a particular reaction were calculated by minimizing or maximizing the fluxes of central metabolic reactions with constraints [25]. All simulations were conducted in Python environment with Gurobi Optimizer 6.0 and GurobiPy package (Gurobi Optimization, Inc., Houston, TX). Reading, writing, and manipulation of the COBRA-compliant SBML files were implemented using COBRApy [26].

3 Results

3.1 Genetic variations in the hyper ABE producing *C. acetobutylicum* BKM19

To identify the genetic variations in the hyper ABE producing BKM19 strain, the amplified DNA library of the BKM19 was re-sequenced by Illumina Miseq with an average length of 51 base pairs. The amplified DNA library of PJC4BK was also re-sequenced as a control under the same condition. In total, 589 Mb and 587 Mb read data were obtained from the chromosomes of the PJC4BK and BKM19 strains, respectively. Also, 37 Mb and 38 Mb read data were obtained from the megaplasmid pSOL1 present in the PJC4BK and BKM19 strains, respectively. These read depths correspond to at least 150 and 190 times those of the chromosome and pSOL1 megaplasmid of the *C. acetobutylicum* ATCC 824, respectively. Read data were mapped using the CLC Genomics Workbench to *C. acetobutylicum* ATCC 824 as the reference sequence (NC_003030.1 for the chromosome and NC_001988.2 for the megaplasmid) to identify the genomic position of the SNVs and InDels.

To our surprise, a total of 155 SNVs, 6 multi-nucleotide variants (MNVs) and 29 InDels

were observed in the parental strain PJC4BK (Supporting information, Fig. S1) compared to the reference wild-type ATCC 824 strain. In a recent publication, a large number of mutations have been reported in the genome of the ATCC 824 COSMIC consortium strain, in which 175 SNVs and 48 InDels were included [27]. Most of the variations identified in PJC4BK are shared with those found in the recently re-sequenced ATCC 824 genome (Fig. 1). However, some variations in genes including CA_C1926, CA_C1989, CA_C2771, CA_C2833, CA_C3079, CA_C3428, and CA_P0053 are present only in the ATCC 824 COSMIC consortium strain, but not in the PJC4BK strain. Also, some variations including CA_C2249, CA_C2367, and CA_C3288 are unique to the PJC4BK strain. In addition, the presence of a 5 kb deletion (position 69715 to 74631, NC_001988) in the COSMIC consortium strain remains intact in the PJC4BK strain. These differences are rather surprising, and suggest the importance of verifying the genome sequences for strains propagated in different laboratory conditions. All these changes were confirmed by Sanger sequencing.

Quality-based variant detection by the CLC Genomics Workbench identified 14 potential mutation points (13 SNVs and 1 deletion) specific to BKM19, while only 1 SNV specific to PJC4BK (Fig. 2). Among the BKM19 specific variants, 13 SNVs were located within the coding regions, while 1 deletion was located in an intergenic region. Except for 1 SNV found on the megaplasmid, all SNVs and InDels were found on the BKM19 chromosome. The PJC4BK specific intergenic variant was detected on chromosome. The sequences of all these mutated loci of around 500 bp were verified by Sanger sequencing (Supporting information, Table SI).

Among the BKM19 specific mutations, 8 of the 14 mutated genes are related to metabolism (CA_C0178, CA_C0879, CA_C1047, CA_C1673, CA_C1780, CA_C1783, CA_C2873, and CA_P0115), 5 of the 14 genes are associated with cellular processes and signaling (CA_C0823, CA_C1518, CA_C1690, CA_C2981, CA_C3315), and one gene is linked to information storage and processing (CA_C_2981). In more detail, the *pilT* gene (CA_C1690) encoding the twitching motility protein is known for cell motility [28]. The *pilT* mutant of *Pseudomonas aeruginosa*, lacks

twitching motility and, forms a dense biofilm [29]. It is interesting to note that we observed the BKM19 strain showed tendency to stick to the fermenter wall during fermentation (data not shown). The *thlA* gene (CA_C2873) encoding the main thiolase is the first enzyme in the fermentative pathway (see the next section for more details). The *htpG* gene (CA_C3315) is one of the genes whose expression responds to butanol stress [30]. Overexpression of HtpG led to an enhanced adaptation to high butanol concentrations in *C. acetobutylicum* [31]. Comparative genome analysis revealed a point mutation in the coding region of CA_C3315, replacing lysine with arginine at the 565^{th} residue, which might contribute to the enhanced solvent production in BKM19.

3.2 Variations of the *thlA* gene in the BKM19 strain

From the comparative genomic analysis, SNV was detected in the *thlA* gene encoding thiolase. Thiolase, catalyzing the condensation of two acetyl-CoAs to acetoacetyl-CoA, is the first enzyme of the butyrate/butanol fermentation pathway. Recently, we reported the crystal structure of the *C. acetobutylicum* thiolase and suggested that it might be responsible for controlling metabolic flux through redox-switch modulation [32, 33]. A point mutation in the coding region of CA_C2873 caused the replacement of valine by alanine at the 5^{th} residue. Interestingly, this amino acid mutation has not yet been reported in previous studies. In order to examine if the mutation in thiolase changes its activity, the mutated thiolase was expressed in *E. coli* BL21(DE3) and purified (Fig. S2). The activity of the mutant thiolase was measured by monitoring the change in acetoacetyl-CoA absorbance at 303 nm. The mutant thiolase ($ThlA^{V5A}$) activity was 815.2 U/mg protein, which was 32% higher than that (617.4 U/mg protein) of the wild-type thiolase (Fig. 3).

To examine the effect of the mutant thiolase on butanol production, the mutant gene, ($thlA^{V5A}$), was cloned into a shuttle vector pTHL1-Cm and transformed into *C. acetobutylicum* ATCC 824 and also into its *thlA*-knockdown mutant TKW-A strain, which was constructed by integration of the mobile group II intron [33]. As a control, the wild-type thiolase ($thlA^{WT}$) gene was

also cloned into the same vector and transformed into the ATCC 824 and TKW-A strains. In batch fermentations, 6.1 g/L of *n*-butanol (a yield of 0.14 g/g glucose) and 0.2 g/L of ethanol were produced in the ATCC 824 overexpressing the $thlA^{V5A}$ gene (Fig. 4B), while 5.8 g/L of butanol (a yield of 0.13 g/g glucose) and 5.9 g/L of ethanol were produced in the TKW-A strain overexpressing the $thlA^{V5A}$ gene (Fig. 5B). In previous studies, *C. acetobutylicum* ATCC 824 overexpressing the $thlA^{WT}$ gene showed that a metabolic shift from the acidogenic to solventogenic phase was hampered during fermentation [33, 34] and showed production of similar amounts of solvents and acids, 4.5 g/L *n*-butanol (a yield of 0.11 g/g glucose), 0.3 g/L of ethanol, 7.0 g/L butyric acid and 5.4 g/L acetic acid (Fig. 4A). In addition, the TKW-A strain overexpressing the $thlA^{WT}$ gene produced 4.0 g/L of butanol (a yield of 0.07 g/g glucose) and 9.5 g/L of ethanol (Fig. 5A). Furthermore, the BKM19 strain overexpressing the $thlA^{V5A}$ gene produced 15.0 g/L of butanol (a yield of 0.20 g/g glucose) and 4.6 g/L of ethanol (Fig. S3B), while 13.5 g/L of butanol (a yield of 0.18 g/g glucose) and 7.1 g/L of ethanol were produced in the BKM19 strain overexpressing the $thlA^{WT}$ gene (Fig. S3A). To investigate differences of metabolic fluxes between mutant and wild-type thiolase in the ATCC 824 strain, *in silico* metabolic flux analysis was performed. The simulation results indicated that metabolic fluxes for butanol production pathway increased in the ATCC 824 strain overexpressing $thlA^{V5A}$ gene compared to the ATCC 824 strain overexpressing $thlA^{WT}$ gene (Fig. S4). Taken together, these results suggest that the flux to butanol production was specifically increased in the strains overexpressing the $thlA^{V5A}$ gene, mainly due to the enhanced thiolase activity.

3.3 Structure-based change of the mutant *thl* gene

Recently, the crystal structure of thiolase from *C. acetobutylicum* revealed that the protein is regulated by a redox-switch regulatory mechanism during the biphasic fermentation [33]. The redox-switch modulation occurs through the formation of a reversible disulfide bond between two

catalytic cysteine residues (Cys88 and Cys378). Structural comparison with other thiolases also revealed that the redox-switch modulation is a unique feature found in the *C. acetobutylicum* thiolase [35]. Compared with other thiolases, the *C. acetobutylicum* thiolase possesses a flexible regulatory determinant region (RDR) that undergoes a large structural change upon formation of the disulfide bond. Moreover, mutant thl$^{V77Q/N153Y/A286K}$, which is regulated by non-redox switch modulation due to a rigid RDR, exhibited a higher enzyme activity than that of the wild-type enzyme and showed enhanced butanol production in batch fermentation [33]. Interestingly, the position of the 5th residue (valine) is located distal from the active site of the enzyme, indicating that this mutation might not have a direct effect on enzyme catalysis (Fig. 6). However, this residue is positioned on the first beta sheet (β1) which directly interacts with the eleventh beta sheet (β11). Because β11 is a secondary structure element comprising the RDR, we speculate that the mutation at the 5th residue might indirectly influence the stability of the RDR, and thus increase the enzymatic activity (Fig. 6).

4 Discussion

A hyper ABE producing BKM19 strain developed by random chemical mutagenesis and screening [17] is a promising clostridial strain for the production of solvents. As one of demonstrations, metabolic engineering of the BKM19 stain allowed production of 27.9 g/L of isopropanol-butanol-ethanol in batch fermentation exhibiting 99% of the total solvents produced as fuel alcohols [36]. Thus, it was of interest to understand the mutations introduced into BKM19 that are responsible for enhanced solvents production. In this study, the BKM19 strain was re-sequenced and 14 potential mutations were identified; 13 SNVs were detected within coding regions, while 1 deletion was located in an intergenic region. Among them, we focused on a thiolase mutation because the enhanced solvent production in the BKM19 strain was due to the reduced formation of fluorocitrate, which is toxic to cells, formed from fluoroacetate [17]. It was hypothesized that the improved

capabilities for solvent production from acetyl-CoA might be due to the mutant thiolase.

In our recent report, the crystal structure of *C. acetobutylicum* thiolase suggested that thiolase activity was mediated by redox-switch regulation during biphasic fermentation. Two catalytic residues, Cys88 and Cys378, in thiolase are involved in the formation of a disulfide bond causing a difference in the mode of conformational changes. The mutant thiolase (V77Q/N153Y/A286K) incapable of forming a disulfide bond exhibited a higher enzyme activity than that of the wild-type enzyme and showed enhanced butanol production in batch fermentation [33]. Furthermore, the optimized thiolase (R133G, H156N, and G222V) identified through screening of a random mutant library showed significantly increased thiolase activity when CoA-SH was added [32]. The mutant site of the thiolase in the BKM19 strain is valine 5 which is in the β1 of the secondary structure. This site is far from the catalytic cysteine loop and the regulatory determinant region proposed in our previous study [33]. However, the mutant thiolase ($thlA^{V5A}$) showed a 15% higher activity than that of wild-type thiolase ($thlA^{WT}$). In addition, the final butanol production was elevated by 26% and 23% in ATCC 824 and TKW-A overexpressing the $thlA^{V5A}$ gene, respectively. The reduced ethanol production in TKW-A overexpressing the $thlA^{V5A}$ gene might be caused by competitive consumption of acetyl-CoA which is consumed by both the thiolase and aldehyde alcohol dehydrogenase (AAD). Due to the enhanced enzyme activity, the carbon flux could be changed from C2 synthesis toward the synthesis of C4 metabolites.

When compared with the ATCC 824 genome, 155 SNVs, 6 MNVs, and 29 InDels were observed in the PJC4BK strain, which is a parental strain of BKM19. When the results were first obtained, we thought that there were too many errors in sequence; thus we repeated the sequencing several times. At the end, however, there have recently been similar next generation sequencing results by several different groups [24, 34] indicating the mistakes present in the original genome sequence of *C. acetobutylicum* due to the use of a less sophisticated sequencing technology at that time [38].

In conclusion, we report identification and analysis of all the mutations in the genome of a hyper ABE producing *C. acetobutylicum* BKM19 strain responsible for increased solvents production. Among the BKM19 specific variants, the mutant *thlA* gene was found to be one of the mutations responsible for enhanced butanol production, which was also proven by metabolic engineering studies. Mutations other than the mutation in the *thlA* gene identified in this study require further studies to better understand the mechanisms behind the enhanced solvents production by the BKM19 strain.

Acknowledgements

This work was supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) and by the Intelligent Synthetic Biology Center of Global Frontier Project (2011-0031963 and 2011-0031957) by the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea. Further support by GS Caltex is appreciated.

Conflict of interest

Authors declare that they have conflict of interest as the BKM19 strain is patented and is of commercial interest.

5 Reference

[1] Carlsson, A.S., Yilmaz, J.L., Green, A.G., Stymne, S., et al., Replacing fossil oil with fresh oil - with what and for what? *Eur J Lipid Sci Technol* 2011, *113*, 812-831.

[2] Durre, P., Biobutanol: an attractive biofuel. *Biotechnol J* 2007, *2*, 1525-1534.

[3] Jang, Y.S., Malaviya, A., Cho, C., Lee, J., et al., Butanol production from renewable biomass by clostridia. *Bioresour Technol* 2012, *123*, 653-663.

[4] Durre, P., Fermentative butanol production: bulk chemical and biofuel. *Ann N Y Acad Sci* 2008, *1125*, 353-362.

[5] Mainguet, S.E. and Liao, J.C., Bioengineering of microorganisms for C(3) to C(5) alcohols production. *Biotechnol J* 2010, *5*, 1297-1308.

[6] Peralta-Yahya, P.P. and Keasling, J.D., Advanced biofuel production in microbes. *Biotechnol J* 2010, *5*, 147-162.

[7] Jones, D.T. and Woods, D.R., Acetone-butanol fermentation revisited. *Microbiol Rev.* 1986, *50*, 484-524.

[8] Moon, H.G., Jang, Y.S., Cho, C., Lee, J., et al., One hundred years of clostridial butanol fermentation. *FEMS Microbiol Lett* 2016, *363*.

[9] Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., et al., Fermentative butanol production by clostridia. *Biotechnol Bioeng* 2008, *101*, 209-228.

[10] Cho, C., Jang, Y.S., Moon, H.G., Lee, J., et al., Metabolic engineering of clostridia for the production of chemicals. *Biofuels Bioprod Biorefin* 2014, *9*, 211-225.

[11] Papoutsakis, E.T., Engineering solventogenic clostridia. *Curr Opin Biotechnol* 2008, *19*, 420-429.

[12] Jang, Y.S., Lee, J.Y., Lee, J., Park, J.H., et al., Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *mBio* 2012, *3*, e00314-12.

[13] Annous, B.A. and Blaschek, H.P., Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Appl Environ Microbiol* 1991, *57*, 2544-2548.

[14] Cooksley, C.M., Zhang, Y., Wang, H., Redl, S., et al., Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. *Metab Eng* 2012, *14*, 630-641.

[15] Tomas, C.A., Welker, N.E. and Papoutsakis, E.T., Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Appl Environ Microbiol* 2003, *69*, 4951-4965.

[16] Hu, S., Zheng, H., Gu, Y., Zhao, J., et al., Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. *BMC Genomics* 2011, *12*, 93.

[17] Jang, Y.S., Malaviya, A. and Lee, S.Y., Acetone-butanol-ethanol production with high productivity using *Clostridium acetobutylicum* BKM19. *Biotechnol Bioeng.* 2013, *110*, 1646-1653.

[18] Mao, S., Luo, Y., Zhang, T., Li, J., et al., Proteome reference map and comparative proteomic analysis between a wild type *Clostridium acetobutylicum* DSM 1731 and its mutant with enhanced butanol tolerance and butanol yield. *J Proteome Res* 2010, *9*, 3046-3061.

[19] Yang, S.T. and Zhao, J., *Adaptive engineering of Clostridium for increased butanol production*. US Patent 8450093, 2013.

[20] Green, E.M., Boynton, Z.L., Harris, L.M., Rudolph, F.B., et al., Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology* 1996, *142 (Pt 8)*, 2079-2086.

[21] Jang, Y.S., Malaviya, A., Lee, J., Im, J.A., et al., Metabolic engineering of *Clostridium acetobutylicum* for the enhanced production of isopropanol-butanol-ethanol fuel mixture. *Biotechnol Prog* 2013, *29*, 1083-1088.

[22] Mermelstein, L.D. and Papoutsakis, E.T., *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon

transformation of *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* 1993, *59*, 1077-1081.

[23] Sliwkowski, M.X. and Hartmanis, M.G., Simultaneous single-step purification of thiolase and NADP-dependent 3-hydroxybutyryl-CoA dehydrogenase from *Clostridium kluyveri*. *Anal Biochem* 1984, *141*, 344-347.

[24] Dash, S., Mueller, T.J., Venkataramanan, K.P., Papoutsakis, E.T., et al., Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. *Biotechnol Biofuels* 2014, 7, 144.

[25] Reed, J.L. and Palsson, B.O., Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res* 2004, 14, 1797-1805.

[26] Ebrahim, A., Lerman, J.A., Palsson, B.O. and Hyduke, D.R., COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* 2013, 7, 74.

[27] Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S.T., et al., Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol Biofuels* 2016, *9*, 4.

[28] Okamoto, S. and Ohmori, M., The cyanobacterial PilT protein responsible for cell motility and transformation hydrolyzes ATP. *Plant Cell Physiol* 2002, *43*, 1127-1136.

[29] Chiang, P. and Burrows, L.L., Biofilm formation by hyperpiliated mutants of *Pseudomonas aeruginosa*. *J Bacteriol* 2003, *185*, 2374-2378.

[30] Tomas, C.A., Beamish, J. and Papoutsakis, E.T., Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *J Bacteriol* 2004, *186*, 2006-2018.

[31] Mann, M.S., Dragovic, Z., Schirrmacher, G. and Lutke-Eversloh, T., Over-expression of stress protein-encoding genes helps *Clostridium acetobutylicum* to rapidly adapt to butanol stress. *Biotechnol Lett* 2012, *34*, 1643-1649.

[32] Mann, M.S. and Lutke-Eversloh, T., Thiolase engineering for enhanced butanol production in *Clostridium acetobutylicum*. *Biotechnol Bioeng.* 2013, *110*, 887-897.

[33] Kim, S., Jang, Y.S., Ha, S.C., Ahn, J.W., et al., Redox-switch regulatory mechanism of thiolase from *Clostridium acetobutylicum*. *Nat Commun* 2015, *6*, 8410.

[34] Sillers, R., Al-Hinai, M.A. and Papoutsakis, E.T., Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in *Clostridium acetobutylicum* fermentations. *Biotechnol Bioeng.* 2009, *102*, 38-49.

[35] Kim, E.J., Son, H.F., Kim, S., Ahn, J.W., et al., Crystal structure and biochemical characterization of beta-keto thiolase B from polyhydroxyalkanoate-producing bacterium *Ralstonia eutropha* H16. *Biochem Biophys Res Commun* 2014, *444*, 365-369.

[36] Jang, Y.S., Woo, H.M., Im, J.A., Kim, I.H., et al., Metabolic engineering of *Clostridium acetobutylicum* for enhanced production of butyric acid. *Appl Microbiol Biotechnol* 2013, *97*, 9355-9363.

[37] Bao, G., Wang, R., Zhu, Y., Dong, H., et al., Complete genome sequence of *Clostridium acetobutylicum* DSM 1731, a solvent-producing strain with multireplicon genome architecture. *J Bacteriol* 2011, *193*, 5007-5008.

[38] Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., et al., Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* 2001, *183*, 4823-4838.

[39] Jang, Y.S., Malaviya, A. and Lee, S.Y., Acetone-butanol-ethanol production with high productivity using *Clostridium acetobutylicum* BKM19. *Biotechnol Bioeng* 2013, *110*, 1646-1653.

[40] Mermelstein, L.D. and Papoutsakis, E.T., *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage φ3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* 1993, *59*, 1077-1081.

Table I. Strains, plasmids, and primers used in this study

Strains/plasmids /primers	Relevant characteristics[a]	Reference or Source
Strains		
C. acetobutylicum ATCC 824	Wild-type	KBRC[b]
C. acetobutylicum PJC4BK	*buk::erm*	[20]
C. acetobutylicum BKM19	Variant of *C. acetobutylicum* PJC4BK	[39]
C. acetobutylicum TKW-A	*thlA::intron*, *thlA* mutant of *C. acetobutylicum* wild-type strain (ATCC 824)	[33]
E. coli TOP10	Host for cloning	Invitrogen[c]
E. coli BL21(DE3)	Host for protein expression	
Plasmids		
pAN1	Cm^r, *Φ3T I* gene, and *p15A* origin	[40]
pTHL1-Cm	Ap^r, MLS^r, Th^r, pTHL1 derivative	[21]
pTHL1-ThlA	pTHL1-Cm derivative containing *thlA*	This study
pTHL-V5A	pTHL1-Cm derivative containing mutant *thlA*	This study
pET-30a(+)	Expression vector; Km^r	Novagen[d]
pET-ThlA	pET-30a(+) derivative contacting *thlA*	This study
pET-V5A	pET-30a(+) derivative contacting mutant *thlA*	This study
Primers		
primer 1	ATAT**CTGCAG**ATGAAAGAAGTTGTAATAGC	This study
primer 2	ATAT**CTGCAG**ATGAAAGAAGTTGCAATAGC	This study
primer 3	ATAT**CCCGGG**ATTCAATTTACGCCTAGTAC	This study
primer 4	ATAT**CCATGG**ATGAAAGAAGTTGTAATAGC	This study
primer 5	ATAT**CCATGG**ATGAAAGAAGTTGCAATAGC	This study
primer 6	ATAT**CTCGAG**GCACTTTTCTAGCAATATTG	This study

[a] Cm^r, chloramphenicol/thiamphenicol resistance gene; MLS^r, erythromycin-resistance gene; Ap^r, ampicillin-resistance geneS

[b] KBRC, Korea Research Institute of Bioscience and Biotechnology Biological Resource Center, Daejeon, Republic of Korea

[c] Invitrogen, Invitrogen Corporation, Carlsbad, CA, USA

[d] Novagen, Madison, WI, USA

Figure legends

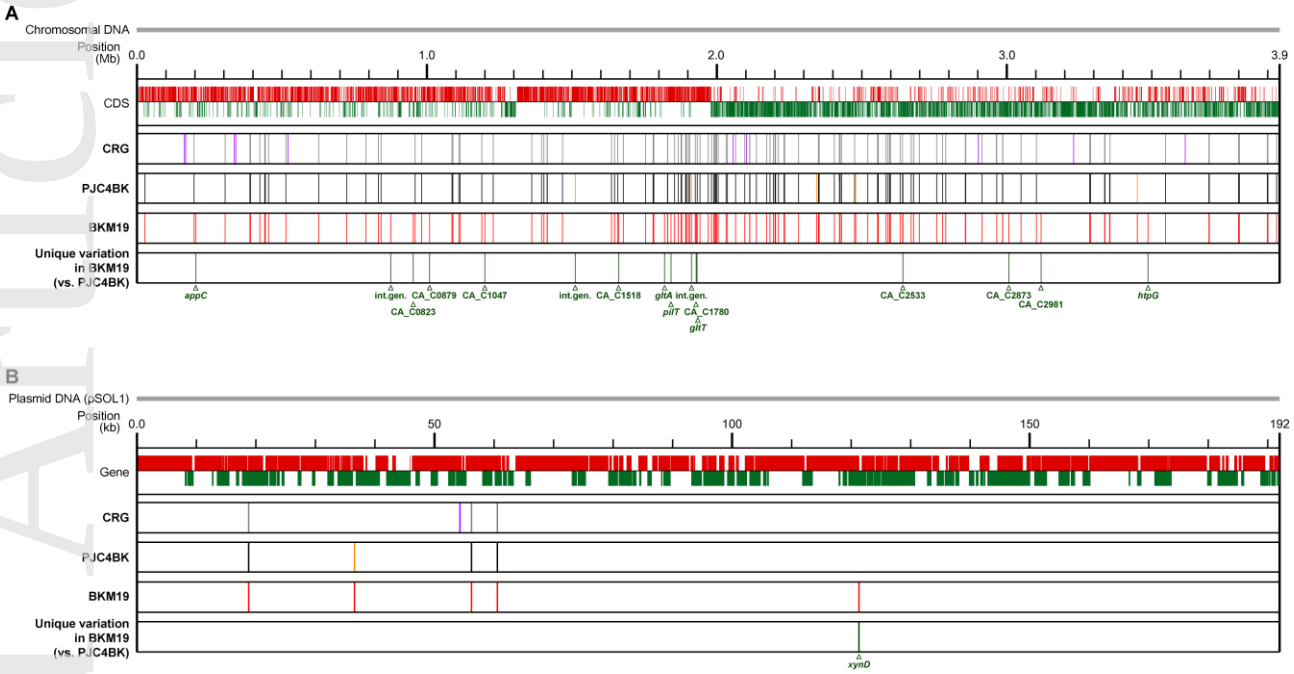
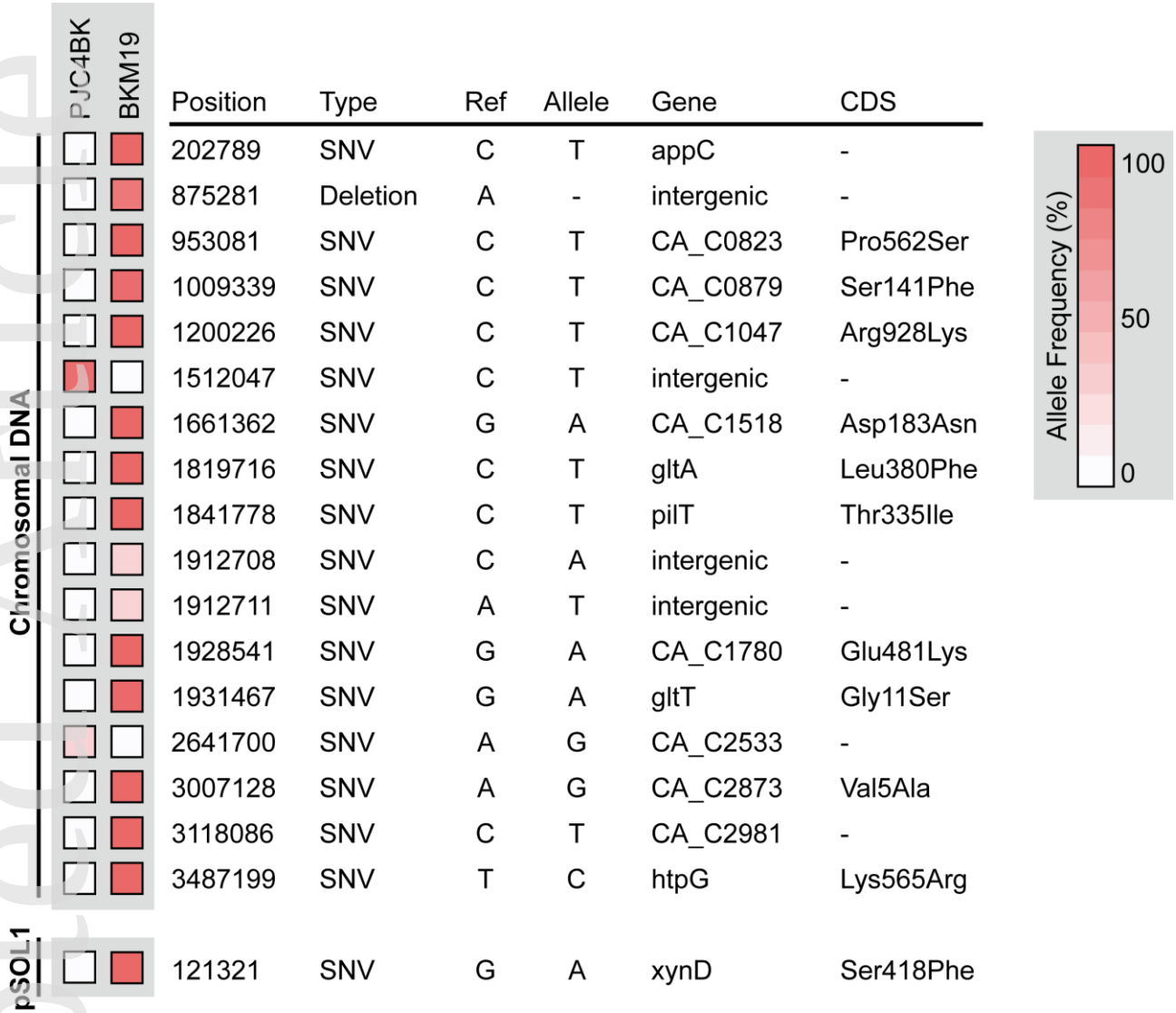


Figure 1. Comparison of the genome of COSMIC consortium strain (CRG), and PJC4BK, and BKM19 strain. The tracks indicate variants compared with ATCC 824. Variation of genes presented in only CRG are indicated as purple lines, and variations showed only in PJC4BK are indicated as orange lines. Unique variations in the BKM19 compared with PJC4BK are indicated as green lines.



Position	Type	Ref	Allele	Gene	CDS
202789	SNV	C	T	appC	-
875281	Deletion	A	-	intergenic	-
953081	SNV	C	T	CA_C0823	Pro562Ser
1009339	SNV	C	T	CA_C0879	Ser141Phe
1200226	SNV	C	T	CA_C1047	Arg928Lys
1512047	SNV	C	T	intergenic	-
1661362	SNV	G	A	CA_C1518	Asp183Asn
1819716	SNV	C	T	gltA	Leu380Phe
1841778	SNV	C	T	pilT	Thr335Ile
1912708	SNV	C	A	intergenic	-
1912711	SNV	A	T	intergenic	-
1928541	SNV	G	A	CA_C1780	Glu481Lys
1931467	SNV	G	A	gltT	Gly11Ser
2641700	SNV	A	G	CA_C2533	-
3007128	SNV	A	G	CA_C2873	Val5Ala
3118086	SNV	C	T	CA_C2981	-
3487199	SNV	T	C	htpG	Lys565Arg
121321	SNV	G	A	xynD	Ser418Phe

Figure 2. The BKM19 Specific variants compared to its parental strain PJC4BK. Heat map showing the allele frequency in PJC4BK and BKM19 strain.

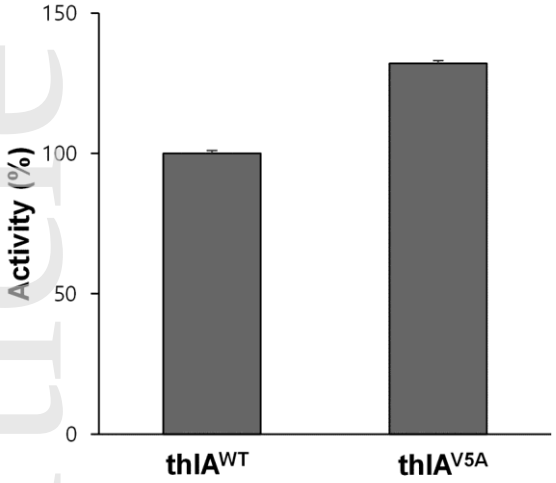


Figure 3. Thiolase activity of the wild-type thiolase (thlA^{WT}) and mutant thiolase (thlA^{V5A}). The thlA^{V5A} mutant shows 32% increase compared with the wild type enzyme. Error bars represent standard deviation among three independently repeated experiments.

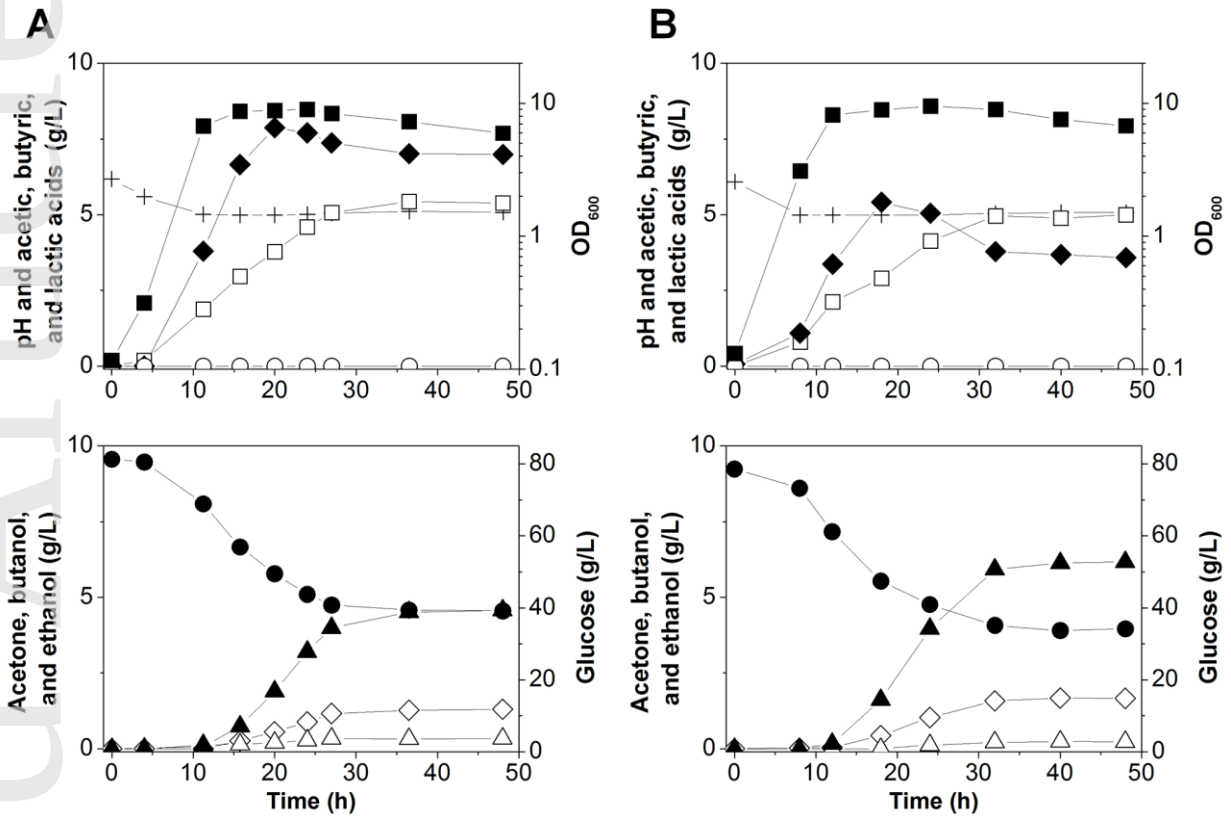


Figure 4. Time profiles of batch fermentation of the *C. acetobutylicum* ATCC 824 strain overexpressing the *thlA*^{WT} gene (A) and the *thlA*^{V5A} gene (B). Symbols are: ■, OD600; □, acetic acid; ○, lactic acid; ◆, butyric acid; +, pH; ●, glucose; ◇, acetone; △, ethanol; ▲, butanol. Batch fermentations were conducted at least in duplicates for reproducibility check, and representative profiles are presented.

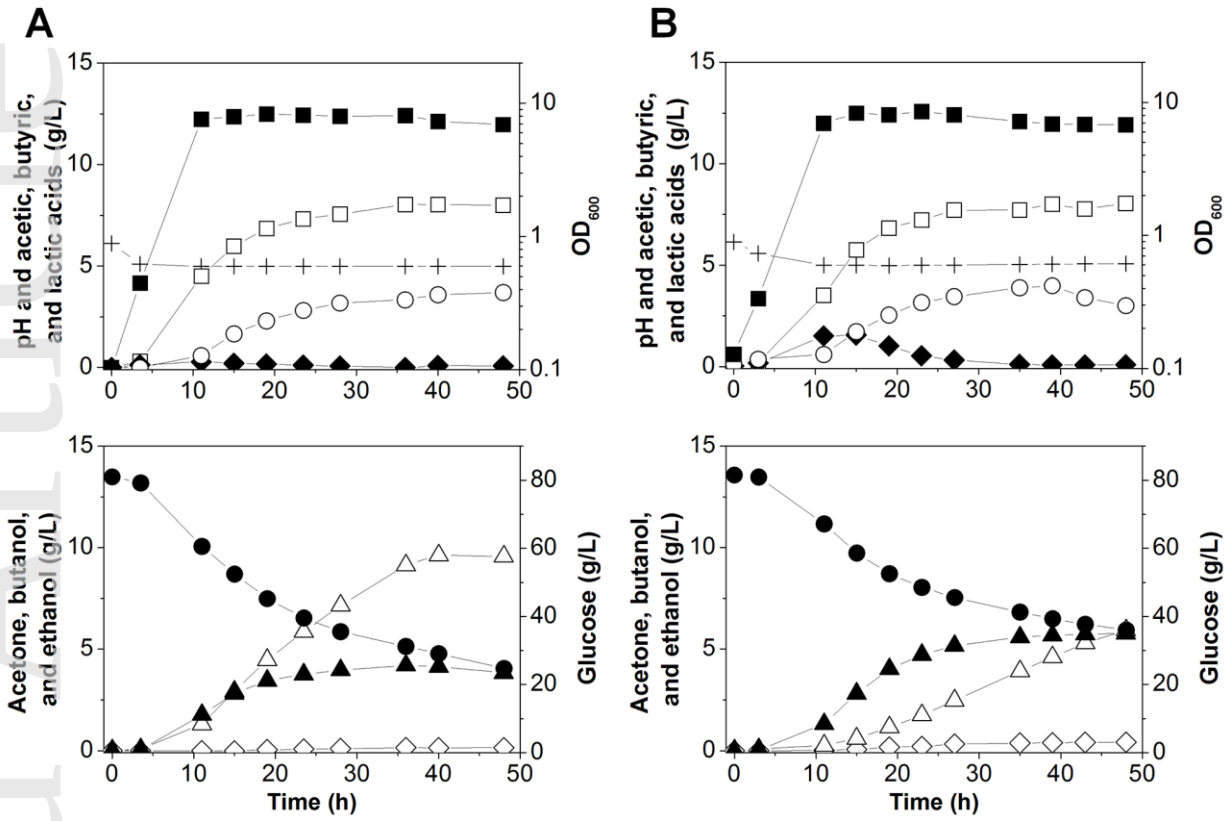


Figure 5. Time profiles of batch fermentation of the *C. acetobutylicum* TKW-A strain overexpressing the *thlA*^{WT} gene (A) the *thlA*^{V5A} gene (B). Symbols are: ■, OD600; □, acetic acid; ○, lactic acid; ◆, butyric acid; +, pH; ●, glucose; ◇, acetone; △, ethanol; ▲, butanol. Batch fermentations were conducted at least in duplicates for reproducibility check, and representative profiles are presented.

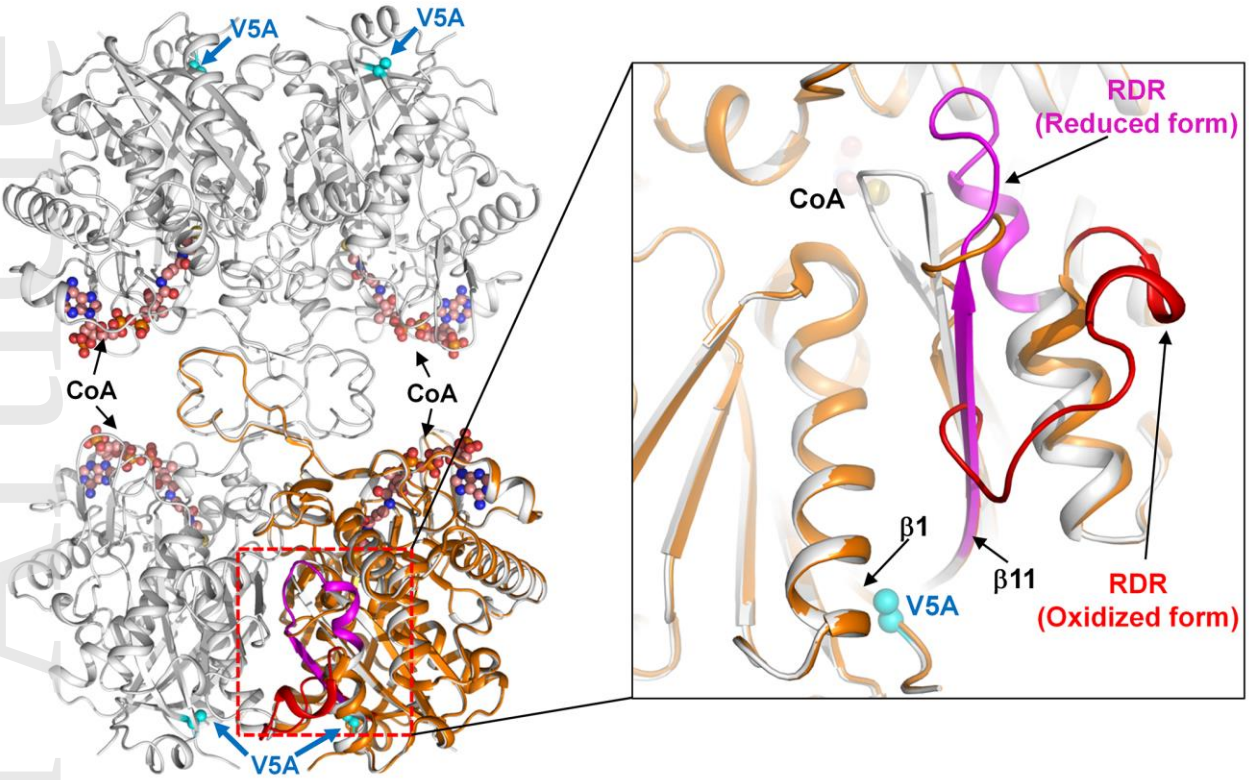


Figure 6. **Position of V5A residue in *C. acetobutylicum* thiolase.** Tetrameric structure of *C. acetobutylicum* thiolase in reduced form is shown as a cartoon diagram with a grey color (left). A monomer of thiolase in oxidized form (orange color) is superposed with one subunit of tetramer in reduced form. The bound CoA molecules are shown as sphere models. Four residues of V5A in tetramer are shown as sphere models with a cyan color and indicated with arrows. Closed-up view of RDR and V5A (right). RDRs in reduced form and in oxidized form are distinguished magenta and red colors, respectively. Structural analysis is performed using PyMOL.