Synthetic and Systems Biotechnology 6 (2021) 384–395

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Inhibitor tolerance and bioethanol fermentability of levoglucosan-utilizing *Escherichia coli* were enhanced by overexpression of stress-responsive gene *ycfR*: The proteomics-guided metabolic engineering

Dongdong Chang [a], Zia Ul Islam [a,b], Junfang Zheng [c], Jie Zhao [a], Xiaoyong Cui [d], Zhisheng Yu [a,e,*]

[a] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, PR China
[b] Department of Chemistry, Mississippi State University, Mississippi State, MS 39762, USA
[c] Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Capital Medical University, Beijing, 100069, PR China
[d] Yanshan Earth Critical Zone and Surface Fluxes Research Station, Chinese Academy of Sciences, No. 380 Huaibei Town, Huairou District, Beijing, 101408, PR China
[e] RCEES-IMCAS-UCAS Joint-Lab of Microbial Technology for Environmental Science, Beijing, 100085, PR China

ARTICLE INFO

Keywords:
Lignocellulosic biomass
Levoglucosan
Inhibitor
Proteomics
Gene overexpression
Bioethanol

ABSTRACT

Pretreatment of lignocellulosic biomass is crucial for the release of biofermentable sugars for biofuels production, which could greatly alleviate the burgeoning environment and energy crisis caused by the massive usage of traditional fossil fuels. Pyrolysis is a cost-saving pretreatment process that can readily decompose biomass into levoglucosan, a promising anhydrosugar; however, many undesired toxic compounds inhibitory to downstream microbial fermentation are also generated during the pyrolysis, immensely impeding the bioconversion of levoglucosan-containing pyrolysate. Here, we took the first insight into the proteomic responses of a levoglucosan-utilizing and ethanol-producing *Escherichia coli* to three representative biomass-derived inhibitors, identifying large amounts of differentially expressed proteins (DEPs) that could guide the downstream metabolic engineering for the development of inhibitor-resistant strains. Fifteen up- and eight down-regulated DEPs were further identified as the biomarker stress-responsive proteins candidate for cellular tolerance to multiple inhibitors. Among these biomarker proteins, YcfR exhibiting the highest expression fold-change level was chosen as the target of overexpression to validate proteomics results and develop robust strains with enhanced inhibitor tolerance and fermentation performance. Finally, based on four plasmid-borne genes encoding the levoglucosan kinase, pyruvate decarboxylase, alcohol dehydrogenase, and protein YcfR, a new recombinant strain *E. coli* LGE-*ycfR* was successfully created, showing much higher acetic acid-, furfural-, and phenol-tolerance levels compared to the control without overexpression of *ycfR*. The specific growth rate, final cell density, ethanol concentration, ethanol productivity, and levoglucosan consumption rate of the recombinant were also remarkably improved. From the proteomics-guided metabolic engineering and phenotypic observations, we for the first time corroborated that YcfR is a stress-induced protein responsive to multiple biomass-derived inhibitors, and also developed an inhibitors-resistant strain that could produce bioethanol from levoglucosan in the presence of inhibitors of relatively high concentration. The newly developed *E. coli* LGE-*ycfR* strain that could eliminate the commonly-used costly detoxicification processes, is of great potential for the *in situ* cost-effective bioethanol production from the biomass-derived pyrolytic substrates.

1. Introduction

Extensive usage of fossil fuels has resulted in a sharp decline in fossil reserves and increased emission of greenhouse gases over the past few decades. The energy and environmental concerns arising from the highly energy-dependent society have thus propelled worldwide research towards developing alternative, sustainable, and renewable energy sources [1]. Lignocellulosic feedstock is the most scalable alternative energy source for second-generation biofuel production [2], which could meet our future energy demand in a large part and

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, PR China.
E-mail address: yuzs@ucas.ac.cn (Z. Yu).

https://doi.org/10.1016/j.synbio.2021.11.003
Received 21 June 2021; Received in revised form 28 August 2021; Accepted 8 November 2021

contribute much to the global goals of carbon emission peaking and neutralization. However, unlike crop-based feedstock, lignocellulose is not readily available for downstream utilization due to its recalcitrant nature. Therefore, complex pretreatment processes involving enzymatic hydrolysis, acid-hydrolysis, or pyrolytic cracking are necessary to decompose lignocellulosic biomass into fermentable intermediates.

Among all the pretreatment processes involved in biofuel production from lignocellulosic biomass, pyrolysis is one of the competitive pathways for biofuel production in cost-effectiveness and ease of operation [3]. Pyrolysis of biomass yields bio-oil containing various intermediate compounds such as furans, organic acids, aromatic compounds and anhydrosugars [4]; therein, levoglucosan is the most abundant and promising anhydrosugar substrate for microbial utilization [5]. Depending on the raw material used, the sugars and inhibitors in lignocellulosic biomass vary in a wide range, with levoglucosan between 6 and 150 g/L, acetic acid between 1 and 99 g/L, furfural between 1 and 11 g/L, and phenol between 1 and 7 g/L are most common [5,6]. However, naturally existing microorganisms cannot directly ferment levoglucosan to biofuels like bioethanol at high yield and productivity [5]. Layton et al. [7] and us [8] has genetically engineered *E. coli* strain to directly ferment levoglucosan to ethanol, however, the furans, organic acids, and aromatic compounds present in the non-detoxified bio-oil inhibited the growth of the bacteria [9,10]. Although many physical, chemical, and hybrid detoxification processes have been developed to remove the bio-oil inhibitors [9,10], the overall costs are considerably elevated by them. Hence, developing robust microbial strains capable of *in situ* bio-detoxification and fermentation of the inhibitors-containing bio-oil is highly desirable.

Previous studies on engineering *E. coli* for inhibitor resistance mainly paid attentions to a single inhibitor [11–13]. Considering that in the bio-oil media there are a variety of inhibitors exerting a synergistic effect on microbial growth and fermentation, robust engineered strain capable of resisting multiple inhibitors need to be developed to overcome the limitations of strains resistant to a single inhibitor. For the development of genetic engineering strains, reveal of the genes related to the resistance of multiple inhibitors is crucial. Proteomics technique, as a powerful tool for evaluating how cells respond to the inhibitors at the protein level, could reveal how genes corresponding to the differentially expressed proteins (DEPs) could be modified to enhance the cells' inhibitor tolerance, rendering the downstream metabolic engineering manipulation feasible [14–18].

In this study, we firstly investigated the cellular responses of a levoglucosan-utilizing and ethanol-producing *E. coli* strain at proteomics level by culturing the bacteria in the presence of furfural, acetic acid, and phenol, both individually and collectively, to identify the underlying proteins (genes) related to inhibitor-resistance. Then based on the proteomics results, we further engineered a strain capable of producing ethanol from levoglucosan in the presence of biomass-derived inhibitors of relatively high levels. The schematic representation of our experimental design and analysis is shown in Fig. 1. To our knowledge, this is the first study investigating the cellular responses of the levoglucosan-utilizing ethanologenic *E. coli* strain to the inhibitory conditions, to reveal the stress-responsive proteins of the strain and consequently develop better microbial strains resisting a wide range of biomass-derived inhibitors.

2. Material and methods

2.1. Microorganism and culture conditions

The levoglucosan-utilizing and ethanol-producing *E. coli* LGE2 previously engineered by us [8] and the lab-collected *E. coli* BL21 (DE3) were used in this study. Liquid and solid LB medium was used for culturing *E. coli* LGE2. The first-grade seed culture obtained by

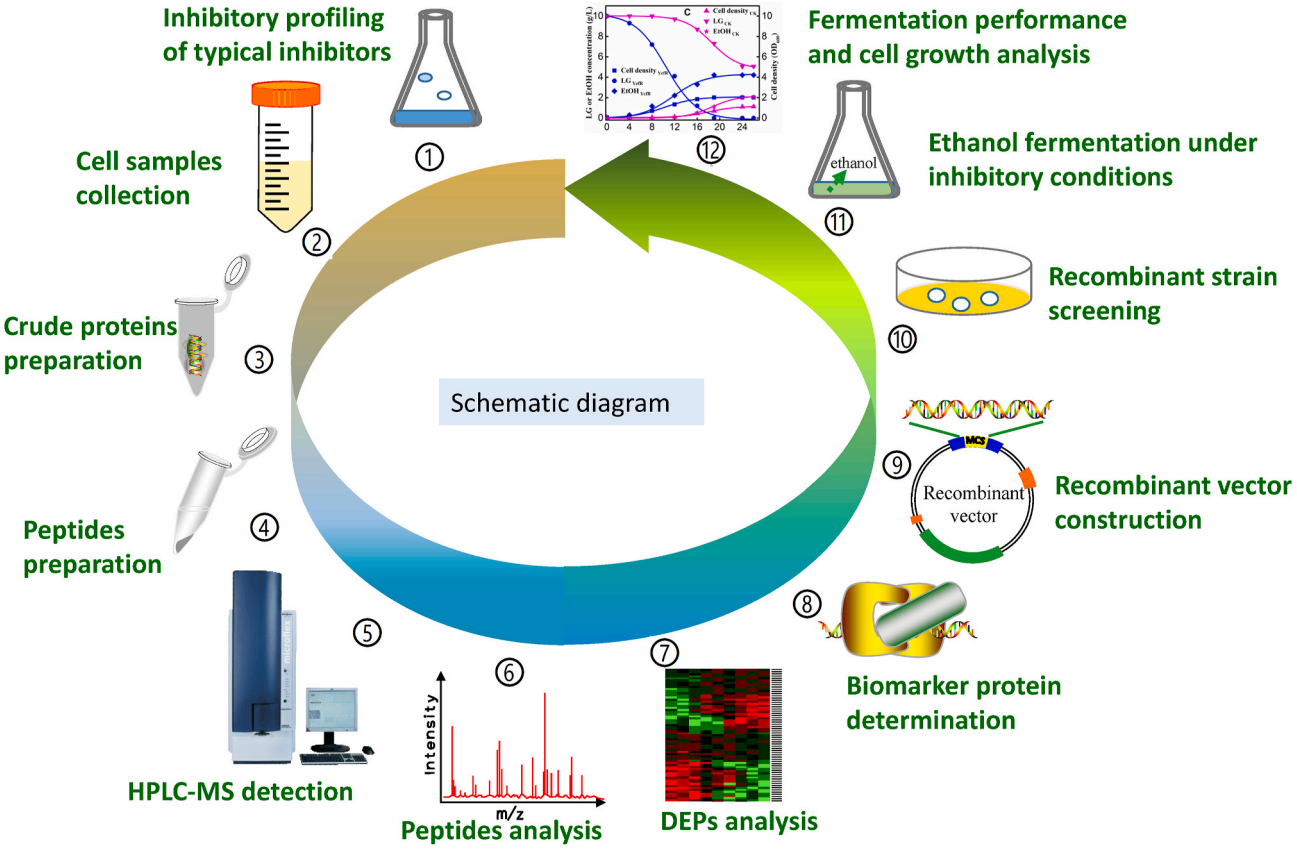


Fig. 1. Schematic diagram of the sampling and analysis experiments adopted in this work.

inoculating a single colony into LB medium was further inoculated (1% v/v) into levoglucosan-containing M9 minimal medium (7.10 g/L Na_2HPO_4, 3.00 g/L KH_2PO_4, 0.50 g/L NaCl, 1.00 g/L NH_4Cl, 0.49 g/L $MgSO_4$, 14.70 mg/L $CaCl_2$, and 10.00 g/L levoglucosan) to obtain the second-grade seed culture. Finally, after incubated anaerobically at 37 °C and 150 rpm overnight, the culture with a cell density of ~6.7 × 10^8 cells/ml was used for subsequent experiments. Ampicillin and chloramphenicol with a final concentration of respective 100 μg/mL and 34 μg/mL were added into the media mentioned above unless otherwise mentioned.

2.2. Determination of the inhibitory concentrations of the inhibitors against cell growth

The inhibitory effects of the representative inhibitors (acetic acid, furfural, and phenol) on cell growth were first determined by adding different concentrations of each inhibitor to the autoclaved M9 minimal media and LB media individually before inoculation; the media without inhibitors were used as control. Cells were grown and sampled at the same time intervals. Each cell sample's optical density was measured using a UV spectrophotometer (Unico Instrument Co., Ltd., Shanghai, China). Changes in OD_{600} values were detected to show the effects of different inhibitors on cell growth. Increasing concentrations of each inhibitor with an increment of 0.05 g/L were added to the culture media to determine each inhibitor's minimum inhibitory concentration (MIC) at respective 24-h and 48-h time points. Then *E. coli* was also grown in the presence of combinations of all the inhibitors containing 0.5%–100% of the MIC of each inhibitor. Three biological replicates were set for each experiment. Based on the results obtained, each inhibitor's appropriate concentration and its combination were used in the next steps to prepare samples and perform proteomics analysis.

2.3. Sample preparation and protein extraction

The *E. coli* cells exposed to different inhibitors were harvested at the logarithmic growth phase, with an OD_{600} value of 0.68 ± 0.05 for proteomics analysis. Cells exposed to acetic acid, furfural, phenol, and their combination were named A, F, P, and C, respectively, and cells grown in the absence of inhibitors were named CK. All the experimental data was collected by averaging of triplicate. The harvested cells were collected, washed, and finally collected by centrifugation. The collected cell pellets were used for protein extraction.

For protein extraction, the collected cells were first suspended into a lysis buffer (8 M Urea, 1 mM EDTA, 1 mM PMSF, 100 mM Tris-HCl, PH 8.5), followed by ultrasonication on ice for 15 min and centrifugation at 15,000×*g* at 4 °C for 15 min to remove sediment. The proteins were reduced at 35 °C for 60 min with 10 mM dithiothreitol and alkylated with 50 mM iodoacetamide at room temperature for 40 min in the dark. Protein digestion was performed by the FASP (filter aided sample preparation) method with trypsin in 100 mM NH_4HCO_3, and finally the peptide concentrations were measured using a NanoDrop 2000 instrument (Thermo Scientific, Wilmington) at an absorbance of A280 nm.

2.4. High pH reverse phase separation of peptides

The peptide mixture was resolved in buffer A (20 mM ammonium formate in pure water, pH 10.0) and fractionated by high pH reverse phase separation using LC-20AB HPLC system (Shimadzu, Japan), with a 4.6 mm × 150 mm, Gemini-NX 5u C18 110A column (Phenomenex, Guangzhou, China) and a linear gradient starting from 5% buffer B to 80% buffer B in 30 min (buffer B: 20 mM ammonium formate in 100% acetonitrile, pH 10.0). The peptide fractions were then collected and dried in a vacuum concentrator (Christ RVC 2-25, Christ, Germany) for downstream analysis.

2.5. Spectral library generation

Data-dependent acquisition (DDA) analysis was performed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, San Jose, California) equipped with an EASY-nLC 1200 system (Thermo Fisher Scientific, San Jose, California). About 3 μg of peptides with iRT peptides (Biognosys, Schlieren, Switzerland) were loaded onto a 100 μm inner diameter × 2 cm C18 trap column at a maximum pressure of 280 bar with 12 μL solvent A (0.1% formic acid in water), and then separated on a 150 μm ID × 25 cm C18 column (1.9 μm, 120 Å, Dr. Maisch GmbH). The gradient with a flow rate of 600 nL/min was 7%–15% solvent B for 15 min, 15%–30% solvent B for 70 min, 30%–50% solvent B for 25 min, and 90% solvent B for 10 min. Data were acquired with full scans (*m/z* 300–1400) using an Orbitrap mass analyzer at a mass resolution of 60,000 at m/z 200. The top 20 precursor ions were selected for fragmentation in the high energy collision dissociation cell at a normalized collision energy of 28%, and then the fragment ions were transferred into the Orbitrap analyzer operating at a resolution of 15,000 at m/z 200. The automatic gain control (AGC) was set to 3e6 for full MS and 5e4 for MS/MS, with maximum ion injection times of 80 and 100 ms, respectively. Dynamic exclusion was set for 1/2 of peak width (18 s).

2.6. DIA analysis

Data-independent acquisition (DIA) was performed using the same mass spectrometer and LC system to DDA, with the same column, same flow rate, and same buffers. The full scan was set at a resolution of 60,000 over an *m/z* range of 350–1400, followed by DIA scans with 30,000, NCE: 27%, AGC target: 1e6, and maximal injection time: 45 ms. Fifty windows were set for DIA acquisition, ranging from 400 to 1200 *m/z*, using an isolation width of 16 *m/z*.

2.7. Data analysis and bioinformatics analysis

Protein identification and quantification were conducted with the Spectronaut pulsar X 12.0 (Biognosys, Boston) with default settings. First, the DDA raw files were searched in the Spectronaut pulsar against the *E. coli* BL21 (DE3) UniProt database (http://www.uniprot.org/uniprot/) to generate a spectral library using BGS factory settings. Peptides FDR was all set as 1%, and the best 3 to 6 fragments per peptide were chosen for the spectral library. The iRT calibration R^2 was set as 0.8. Next, the DIA data were input into the software for protein quantification. The main parameter iRT regression type was set as local (non-linear) regression. All the results were filtered by a Q value cutoff of 0.01, corresponding to an FDR of 0.01. Density Estimator was performed to estimate p-value, and the peak area was used for quantification. Each peptide contained at least three fragment-ions.

The paired difference test was used to identify DEPs. Proteins with $\log_2FC > 1$ or < -1 (FC, fold change) and p-value <0.05 were selected as DEPs. Functional enrichment of these DEPs was conducted by KEGG, GO, COG, and UniProt analysis. Interaction networks of protein to protein and protein to metabolic pathway were constructed and analyzed using the String v11.0 [19] and BioGRID v4.2.191 database [20].

2.8. Plasmid and recombinant strain construction

The pET-*lgk* vector (Fig. S1A, Supplementary file 1) carrying the *Lipomyces starkeyi*-derived and codon-optimized *lgk* gene (GenBank Accession: MT219959) and pZBC vector (Fig. S1B, Supplementary file 1) carrying the *Zymomonas mobilis*-derived pyruvate decarboxylase gene *pdc* (GenBank Accession: M15393.2) and alcohol dehydrogenase II gene *adh* (GenBank Accession: M15394.1) constructed by us [8] were used. YcfR protein was selected as the target of overexpression owing to its highest FC values among all the DEPs. The *ycfR* gene (GenBank Accession: NF033776) encoding YcfR was PCR amplified from the genome of

E. coli BL21 (DE3) using Phusion Polymerase (New England Biolabs, Ipswich, MA) and a pair of primers-forward primer 5′-3′ CGCGGATCCATGAAAATCAAAACCAC and reverse primer 5′-3′ CATGAATTCT**TAACTTTAAGAAGGAG**TTATTTGTACAGTTCAG. The underlined sequences are restriction enzyme sites, and the bold sequence is the ribosome binding site sequence. The *ycfR* gene was then cloned into the pET-*lgk* vector between the *Bam*HI and *Eco*RI restriction sites upstream of the *lgk* gene. The obtained vector pET-*ycfR-lgk* (Fig. S1C, Supplementary file 1) was sequence verified by Sangon Biotech (Shanghai) Co., Ltd, and then transformed to *E. coli* BL21 (DE3) to obtain the primary recombinant strain *E. coli* (pET-*ycfR-lgk*). The pZBC vector was subsequently transformed into the competent *E. coli* (pET-*ycfR-lgk*) to obtain the final recombinant strain *E. coli* LGE-*ycfR*. Expression vectors were transformed into *E. coli* competent cells by electroporation method [8]. Transformed cells were resuspended in LB media, recovered at 37 °C for 1 h, and then grown for ~11 h on LB plates with the addition of appropriate antibiotics to select the positive transformants.

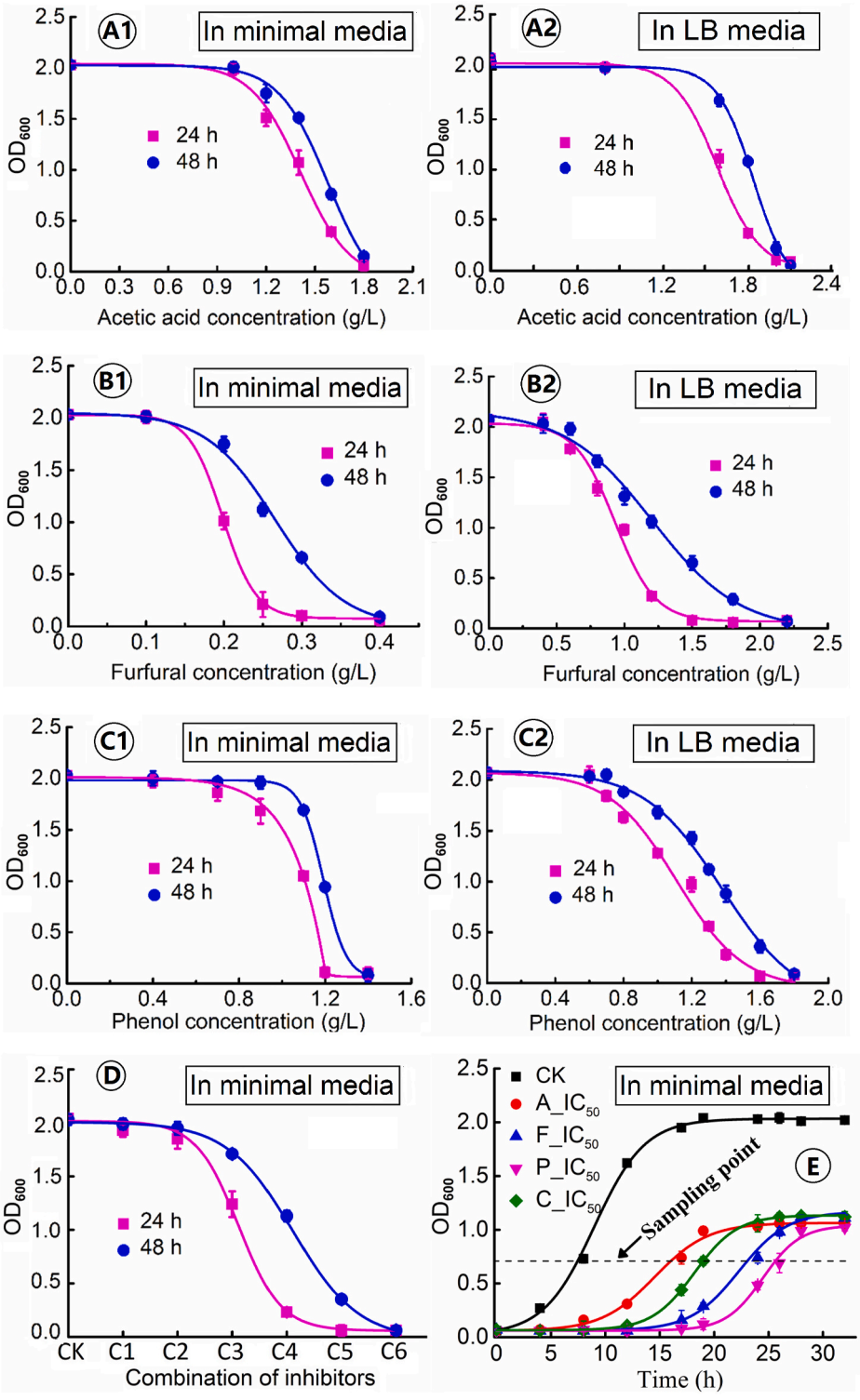


Fig. 2. Inhibitory profiles of individual and combined inhibitors on the cell growth of *Escherichia coli* LGE2. ***Panel A1*** Cells challenged by acetic acid in minimal media. ***Panel A2*** Cells challenged by acetic acid in LB media. ***Panel B1*** Cells challenged by furfural in minimal media. ***Panel* B2** Cells challenged by furfural in LB media. ***Panel C1*** Cells challenged by phenol in minimal media. ***Panel C2*** Cells challenged by phenol in LB media. ***Panel D*** Cells challenged by combined inhibitors in minimal media; the labels CK, C1, C2, C3, C4, C5, and C6 at horizontal axis represent respective combination of 0%, 25%, 35%. 45%, 55%, 60%, and 65% of the individual MIC (minimum inhibitory concentration) of each inhibitor. ***Panel E*** Cells challenged by individual and combined inhibitors at their IC_{50} levels for 48-h incubation in minimal media; A_IC_{50}, F_IC_{50}, P_IC_{50}, and C_IC_{50} denote the concentrations of acetic acid, furfural, phenol, and combined inhibitors inhibiting 50% of the cell growth with 48-h incubation time, respectively; dark dotted line denotes the sampling point (OD_{600} = 0.68 ± 0.05) chosen for further proteomics analysis.

2.9. Inhibitor tolerance test and ethanol fermentation of the engineered strain

Levoglucosan-based M9 minimal media were used to test the recombinant strain's tolerance to acetic acid, furfural, phenol, and their combinations. In contrast, levoglucosan-based M9 minimal media without inhibitors were used as a control. The cultures were incubated at 30 °C and 150 rpm aerobically, with the addition of 0.06 mM isopropyl-β-D-thiogalactopyranoside, 100 μg/mL ampicillin, and 34 μg/mL chloramphenicol. The recombinant's cell growth and fermentation performance were evaluated under the influence of different concentrations of inhibitors to determine the tolerance of the recombinant strain.

2.10. Sampling and analytical method

Sampling and analysis were carried out according to the method described previously [21]. For each sampling point, 5-mL culture media were taken to separate the cells and supernatants. After performing several sequential wash and centrifugation steps, the harvested cells were dried in an oven set at 70 °C to determine the cell dry weights. The specific growth rate μ was calculated by the formula $\mu = \ln(N_2/N_1)/(t_2\text{-}t_1)$, where N_2 and N_1 are the cell dry weights detected in two time points t_2 and t_1, respectively. The clarified supernatants were filtered and collected for HPLC analysis. Analyses of levoglucosan, acetic acid, and ethanol were performed by a high-performance liquid chromatography system (HPLC, LC-20AT, Shimadzu Corp.) described previously [21]. Furfural and phenol were analyzed by HPLC using C18 column (250 mm × 4.6 mm, 5 μm particle size, Thermo Scientific Corp.), SPD-20A UV detector, and mobile phase of 80% methanol aqueous solution. Three replicate samples were evaluated in each case. All reagents used in this study were of analytical grade.

3. Results

3.1. Determination of the inhibitory concentrations of single and combined inhibitors for E. coli LGE2 cells was fundamental for the downstream proteomics analysis

Dose-response curves for cell growth (OD_{600}) of *E. coli* LGE2 under different concentrations of acetic acid, furfural, and phenol in minimal media are shown in Fig. 2. In minimal media, extremely low concentrations of the inhibitors showed no apparent effects on cell growth, while the cell growth decreased when the inhibitor concentrations continuously increased (Fig. 2). Eventually, acetic acid, furfural, and phenol completely blocked the cell growth of *E. coli* LGE2 at their MIC levels, namely IC_{100}, are shown in Table 1. The IC_{50} values indicating the cell growth was inhibited by 50% are also shown in Table 1. As a comparison, *E. coli* LGE2 was parallelly grown in LB media added with different concentrations of the inhibitors. The corresponding cell growth curves and MIC of inhibitors are also presented in Fig. 2A2, 2B2, and 2C2 and Table 1, which show that in LB media the MIC levels of inhibitors was higher than those in minimal media. However, considering minimal medium is of minimal interference with the product of interest, less prone to batch variations, and more relevant to industrial production than nutrient-rich media [22,23], minimal medium was better to be used for subsequent proteomic experiments.

In the sampling strategy, the fundamental thing was to select suitable inhibitor concentrations that could inhibit the cell growth to an appropriate level but not completely block it. In this respect, IC_{50} (for 48-h incubation time) of each inhibitor for M9 minimal medium, that is, 1.45 g/L acetic acid, 0.25 g/L furfural, and 1.20 g/L phenol, were applied to challenge the cells. Moreover, different combinations of each inhibitor's MIC were also used to challenge the *E. coli* LGE2 cells. The combination of 1.00 g/L acetic acid +0.20 g/L furfural +0.80 g/L phenol (~55% of each inhibitor's individual MIC) resulting in a final cell density of 1.13 ± 0.06 (Fig. 2D), was designed as the IC_{50} of the combined inhibitors. For the subsequent proteomics analysis, *E. coli* cells were strictly harvested at the same growth phase (OD_{600} ~0.68) to guarantee the reliability of the final results (Fig. 2E). Fig. 2E also shows that much longer time was required to reach the same density to the control culture when challenging by the inhibitors, among which, phenol was the most toxic and showed the most severe inhibition when each inhibitor was at their corresponding IC_{50}.

3.2. Proteomic profiling of the inhibitors-exposed E. coli LGE2

3.2.1. Overview of quantitative proteomics indicated the distribution of DEPs in different groups had sameness within the difference

The proteomic differences of *E. coli* LGE2 cellular proteins in the presence and absence of inhibitors were analyzed by DIA-based quantitative proteomics, which exhibited a total of 2749 unique proteins corresponding to 268,125 spectra (Sheet S1, Supplementary file 2). Using a cutoff of more than 2-fold change (FC) and a *p*-value less than 0.05, the numbers of DEPs in all the groups were determined and presented in Table S1, Supplementary file 1. The intersection number (both unique and shared protein number) of the DEPs among all the four groups is given in Fig. 3A, while respective 15 and 8 proteins were found collectively up-regulated and down-regulated in all the groups (Fig. 3B and C).

To get a deep insight into these DEPs, distribution of Gene Ontology (GO) categories, enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and Cluster of Orthologous Groups (COG) categories of these DEPs were further analyzed. The GO categories are further classified into three individual ontologies-biological process, cellular component, and molecular function. Most of the DEPs mainly participated in the metabolic and cellular processes, involved in the cell and cell part, and played roles in the catalytic activity and binding function, corresponding to the above three ontologies (Figs. S1A–S4A, Supplementary file 1). KEGG pathway analysis showed that the identified DEPs were mostly involved in metabolism pathway, environmental information processing pathway, and genetic information processing pathway

Table 1
Inhibitory concentrations of inhibitors for *E. coli* LGE2 cultured for respective 24 and 48 h.

Inhibitor	Incubation time (h)	Inhibitory concentration (g/L)[a]				Resultant cell density (OD_{600})[b]			
		For LB media		For minimal media		For LB media		For minimal media	
		IC_{50}	IC_{100} (MIC)	IC_{50}	IC_{100} (MIC)	OD_{IC50}	OD_{IC100}	OD_{IC50}	OD_{IC100}
Acetic acid	24	1.6	2.0	1.4	1.8	1.11	0.09	1.07	0.12
	48	1.8	2.1	1.45	1.8	1.08	0.06	1.03	0.15
Furfural	24	0.8	1.5	0.2	0.3	0.98	0.08	1.01	0.10
	48	1.2	2.2	0.25	0.4	1.06	0.06	1.12	0.09
Phenol	24	1.2	1.6	1.1	1.2	0.97	0.07	1.05	0.11
	48	1.3	1.8	1.2	1.4	1.12	0.09	0.94	0.08

[a] Abbreviations: IC_{50} and IC_{100} denote concentrations inhibiting 50%, and 100% of cell growth, respectively. IC_{100} = MIC, minimal inhibitory concentration.
[b] The cell densities (OD_{600}) in the control for LB media and minimal media were detected as 2.07 and 2.01, respectively.

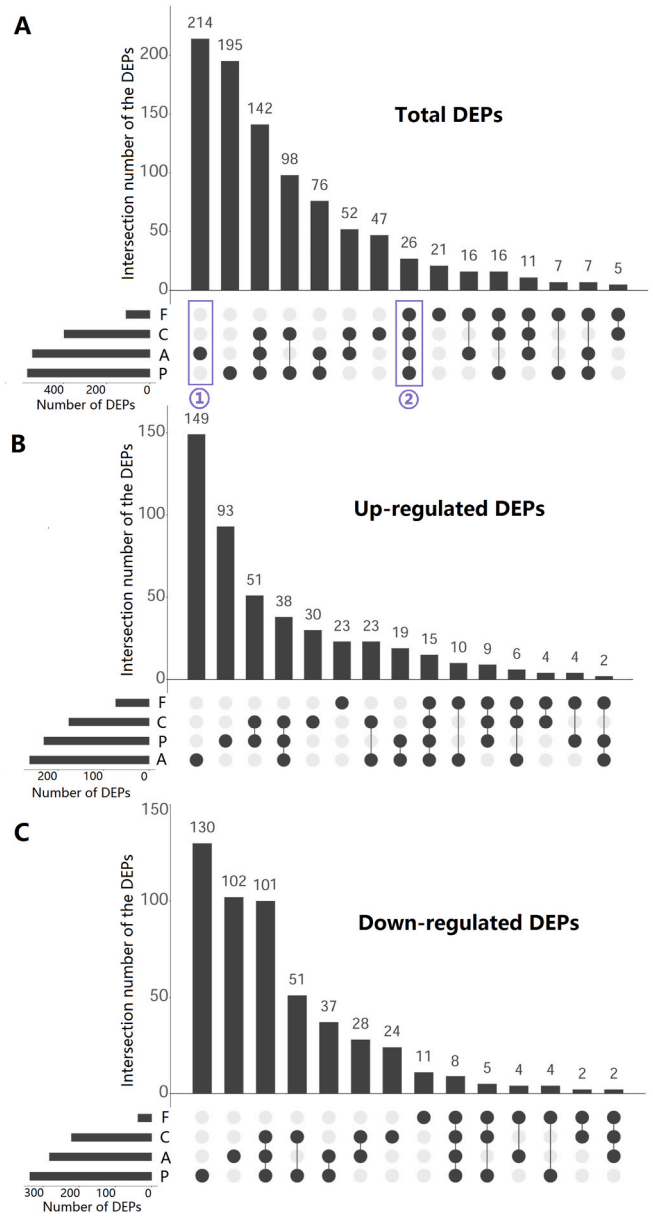


Fig. 3. Numbers of the unique and shared differentially expressed proteins (DEPs) of all the inhibitor-treated groups. ***Panel A*** shows the numbers of unique and shared up-regulated proteins, and ***panel B*** shows the numbers of unique and shared down-regulated proteins. These protein numbers are shown in an UpSet diagram, allowing for a clearer plotting of large data sets compared to the Venn diagram. Dark circles (●) connected with a line indicate that the proteins are only differentially expressed in the corresponding group(s) labelled with ●, but not in the other groups labelled with ●. For example, the label ① denotes the proteins are only differentially expressed in acetic acid treated group, while not differentially expressed in the other groups, that is, these proteins are unique in acetic acid treated group; ② denotes the proteins are differentially expressed in all the groups, that is, these proteins are shared by all the groups. Abbreviations **A**, **F**, **P**, and **C** represent cells treated by acetic acid, furfural, phenol, and combined inhibitors of acetic acid, furfural, and phenol, respectively.

(Figs. S1B–S4B, Supplementary file 1). Functional classification of these DEPs conducted by COG analysis showed that these DEPs were mainly involved in energy production and conversion; DNA replication, recombination, repair, transcription, and RNA translation; amino acid transport and metabolism; and carbohydrate transport and metabolism and so on (Table S2, Supplementary file 1).

3.2.2. Total 23 differentially expressed proteins shared by all the inhibitor-treated groups were identified as stress-responsive biomarker proteins

The 15 up-regulated and 8 down-regulated proteins shared by all the inhibitor-treated groups, which could be assumed to have roles in the inhibitor resistance for the exposed *E. coli* strain, were of our interest. The 15 up-regulated proteins were identified as YcfR (YchN), CyoC, EvgS, Ybl185, Sra, YgcE, PhoH, CheA, RbsA, MdtF, KdsC, YtfE, CyoB, BetT, and CyoA; 8 down-regulated proteins were identified as YebT, SbcB, YfbS, GarL, MurR, AphA, NarJ, and Ea22 (Fig. 4A and Table S2, Supplementary file 1). KEGG and COG analysis showed that these shared DEPs were mainly involved in signal transduction mechanisms, defense mechanisms, and energy production and conversion of *E. coli* LGE2 strain (Fig. 4B), in response to the oxidative stress, DNA damage, unfolded proteins, osmotic stress, and nutrient starvation, etc. Besides, protein-protein interactions analysis (Fig. 4B) of these shared DEPs showed that the most abundant interactions were presented by EvgS that interacted with the DEPs like Ea22, MdtF, CheA, MurR, and YfbS, followed by the protein YfbS that interacted with AphA, KdsC, and EvgS. CyoA, CyoB, and CyoC are also mutually connected with each other; YcfR and YebT, together with YtfE and NarJ, were another two pairs of interactive proteins. The interactions indicate that the DEPs might be neighboring, fusion, co-expression, or experimentally-determined interactive proteins, etc.

Of these 15 up-regulated proteins, YcfR showed the highest average fold change (FC) value, followed by CyoC and EvgS (Fig. 4A, where the FC values were transformed to their respective log-base 2 values). Further, in terms of the FC values of the up-regulated DEPs in a specific inhibitor-treated group (Fig. 4A), YcfR exhibited the maximal FC value in response to the phenol stress, followed by YcfR in response to the stress of combined inhibitors and EvgS in response to the acetic acid stress. Among the 8 down-regulated proteins, Ea22 showed the highest average FC value, followed by NarJ and AphA (Fig. 4A). In terms of the FC values of the down-regulated DEPs in the four different groups, Ea22 in groups A, F, and P exhibited the top three FC values responding to the stresses of acetic acid, furfural, and phenol, respectively. These results indicate that some DEPs like YcfR with high FC values might be important biomarker proteins playing essential roles in helping cells to overcome the inhibitory stresses.

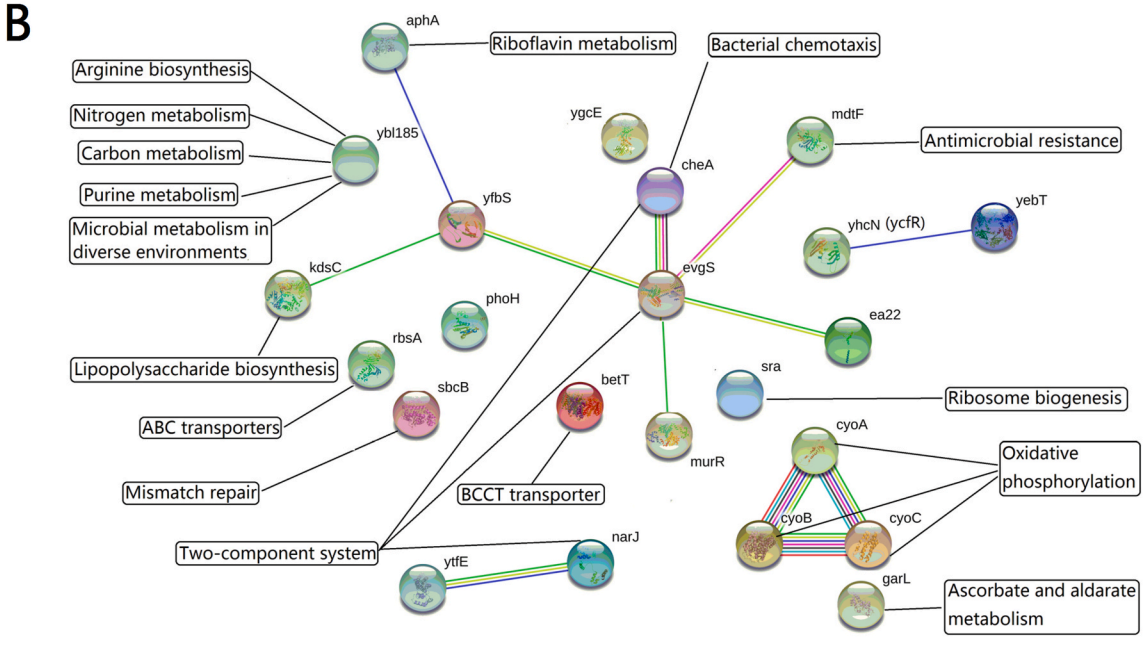
3.3. The engineered strain E. coli LGE-ycfR exhibited improved inhibitor tolerance, cell growth, and bioethanol fermentability when challenged with inhibitors

Among the 23 up- and down-regulated proteins deferentially expressed in all the inhibitor-treated groups (Fig. 4A), the up-regulated protein YcfR showed the highest FC value (144 folds) in response to the phenol stress and also showed the second highest FC value (58 folds) in response to the stresses of the combined inhibitors. Furthermore, in terms of the average FC value, YcfR (56 folds) was also the highest among all the groups. Therefore, we selected the YcfR-encoding gene *ycfR* as the target for overexpression in this work. Subsequently, the engineered strain *E. coli* (pET-*ycfR*-*lgk*, pZBC) harboring three heterologous genes *lgk*, *pdc* and *adh* and one overexpressed gene *ycfR*, named *E. coli* LGE-*ycfR*, was subjected to minimal media containing various concentrations of individual and combined inhibitors. As a result, after a 48-h incubation, *E. coli* LGE-*ycfR* could slightly grow in minimal media supplied with respective 3.8 g/L acetic acid, 1.2 g/L furfural, and 2.3 g/L phenol with a final cell density (OD_{600}) of 0.21, 0.16, and 0.19, respectively, and an increase of 0.1 g/L of these values inhibited the cell growth entirely in each case. Thus, we concluded that the MIC (IC_{100}) values of acetic acid, furfural, and phenol for *E. coli* LGE-*ycfR* were improved to 3.9, 1.3, and 2.4 g/L, respectively, increasing by ~117%, 225%, and 71% compared to the MIC values for the control strain-*E. coli* LGE2.

Moreover, in the minimal media supplied with 1.45 g/L acetic acid, 0.25 g/L furfural, and 1.2 g/L phenol, which are the IC_{50} values of these

A

Protein ID	Protein Name	Gene Name	log2FC A	F	P	C	avg
A0A140N4R7	Protein YcfR	*ycfR*					
A0A140NC92	Cytochrome o ubiquinol oxidase, subunit III	*cyoC*					
A0A140N502	Hpt sensor hybrid histidine kinase	*evgS*					
A0A140NCX7	Carbamate kinase	*ybl185*					
A0A140NCE1	30S ribosomal subunit S22	*sra*					
A0A140N8T7	Carbohydrate kinase FGGY	*ygcE*					
A0A140NAZ5	Protein PhoH	*phoH*					
A0A140N2Z7	Efflux pump membrane transporter	*mdtF*					
A0A140NAQ2	CheA signal transduction histidine kinase	*cheA*					
A0A140N5T4	Ribose import ATP-binding protein RbsA	*rbsA*					
A0A140N5J7	KDO-8-P phosphatases KdsC	*kdsC*					
A0A140NBU6	Iron-sulfur cluster repair protein YtfE	*ytfE*					
A0A140NA88	Cytochrome o ubiquinol oxidase, subunit I	*cyoB*					
A0A140ND74	Choline/carnitine/betaine transporter	*betT*					
A0A140NCS6	Ubiquinol oxidase subunit 2	*cyoA*					
A0A140N6C3	Mammalian cell entry related domain protein	*yebT*					
A0A140N8M8	Exodeoxyribonuclease I	*sbcB*					
A0A140N9Y7	Citrate transporter	*yfbS*					
A0A140N848	5-keto-4-deoxy-D-glucarate aldolase	*garL*					
C5W7D8	HTH-type transcriptional regulator MurR	*murR*					
A0A140ND40	Class B acid phosphatase	*aphA*					
A0A140N9Y6	Molybdenum cofactor assembly chaperone	*narJ*					
A0A140NBC3	Uncharacterized protein	*ea22*					



B

Fig. 4. Heat map (A) and documented interactions (B) of the DEPs shared by all the inhibitor-treated groups. ***Panel A*** Rows are colored by the log2 fold-changes (**FC**) of the proteins in inhibitor-treated groups relative to the corresponding proteins in control. Groups **A**, **F**, **P**, and **C** represent cells treated by acetic acid, furfural, phenol, and combined inhibitors, respectively. **avg** represents the average fold-change of the four groups. The darker the red color, the greater the fold-change of up-regulation. The darker the cyan color, the greater the fold-change of down-regulation. ***Panel B*** Interactions were obtained based on String 11.0 database. A greater number of lines associated with the connection, indicates a greater level of confidence in the association. The network nodes are gene names for the corresponding biomarker proteins. The edges represent the predicted functional associations. An edge may be drawn with up to 7 differently colored lines - these lines represent the existence of the seven types of evidence used in predicting the associations. A red line indicates the presence of fusion evidence; a green line - neighborhood evidence; a blue line - coocurrence evidence; a purple line - experimental evidence; a yellow line - textmining evidence; a light blue line - database evidence; a dark gray line - coexpression evidence. The black boxes connected with the nodes by black lines exhibited the annotated functional categories of the proteins; the others are unannotated proteins in the database.

inhibitors for *E. coli* LGE2, *E. coli* LGE-*ycfR* exhibited much better growth and fermentability than *E. coli* LGE2 without the overexpression of *ycfR* gene in all the cases (Fig. 5). After a maximum of 32-h incubation, the final cell densities of *E. coli* LGE-*ycfR* in each media used (acetic acid added media, furfural added media, and phenol added media) were all ~2.0-folds of the control using *E. coli* LGE2 (Fig. 5 and Table 1). To simplify, the specific growth rate, ethanol concentration, ethanol productivity, and levoglucosan consumption rate listed hereafter for comparisons were all presented in the order of the values in acetic acid added media, furfural added media, and finally phenol added media unless otherwise specified. The specific growth rates of *E. coli* LGE-*ycfR*, as expected, were ~3.0, 2.6 and 2.8-folds of the control, respectively. Furthermore, the ethanol concentration, productivity, and levoglucosan consumption rates obtained by *E. coli* LGE-*ycfR* in all the cases were also remarkably improved compared to the values shown by the control. At the fermentation endpoint, the ethanol concentrations exhibited by *E. coli* LGE-*ycfR* in the presence of these three inhibitors were respective ~2.1, 0.9, and 1.7-folds of the control; the corresponding ethanol productivities respective ~2.6, 2.8, and 2.0-folds; and the levoglucosan consumption rates respective ~2.4, 3.1, and 2-folds. In addition, the

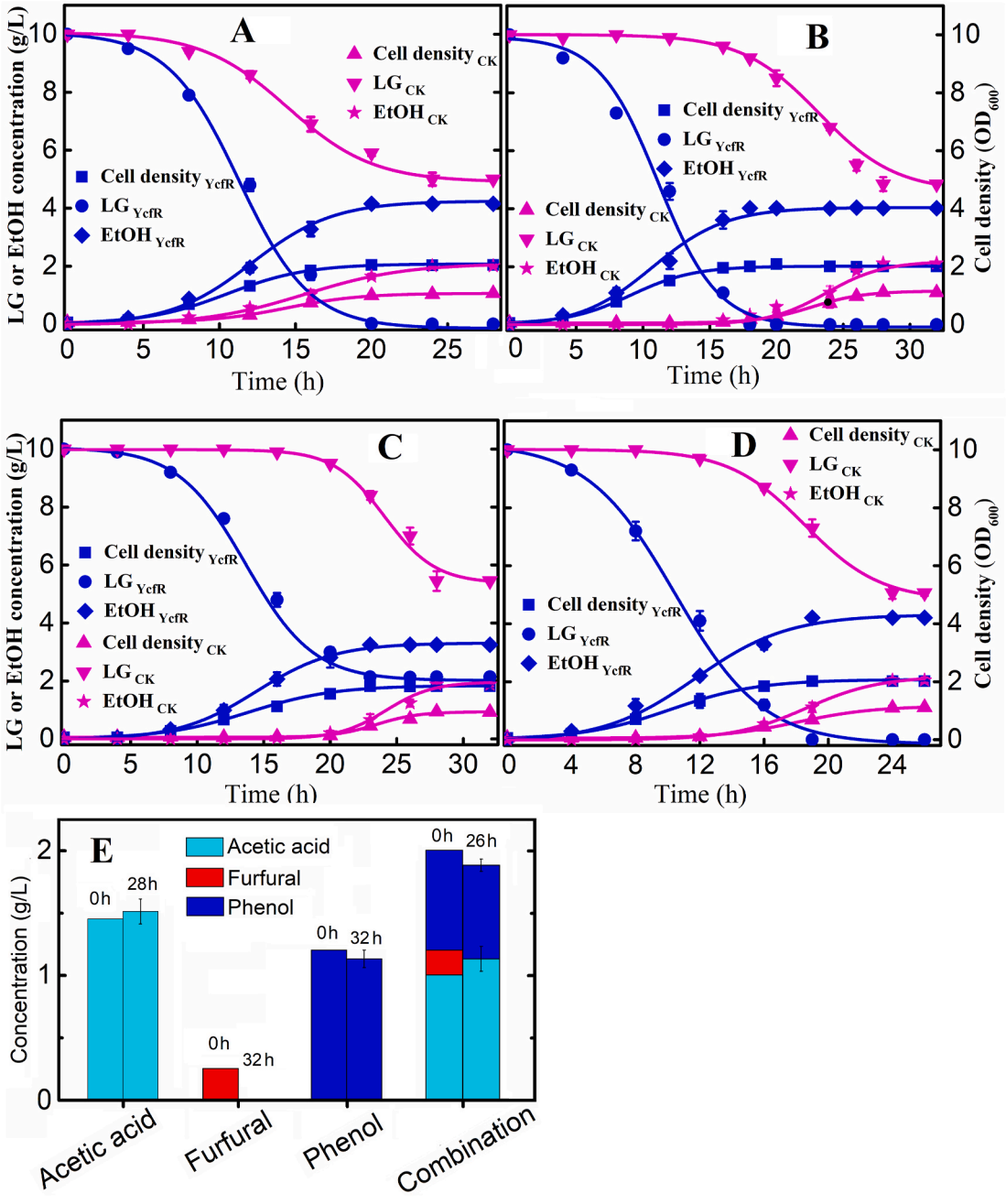


Fig. 5. The time-course changes of cell growth, levoglucosan consumption, ethanol production, and inhibitors concentration in the fermentation tests. *E. coli* LGE-*ycfR* was grown under the IC_{50} of individual and combined inhibitors identified for the non-overexpression strain *E. coli* LGE2 (defined as CK). IC_{50} denotes the inhibitor concentration inhibiting 50% of the cell growth obtained by cultures in the absence of inhibitors. ***Panel A*** cells were grown in the presence of acetic acid; ***Panel B*** cells were grown in the presence of furfural; ***Panel C*** cells were grown in the presence of phenol; ***Panel D*** cells were grown in the presence of combined inhibitors; ***Panel E*** inhibitors' concentrations before and after the fermentation. Cell densities $_{YcfR}$, LG $_{YcfR}$, and EtOH $_{YcfR}$, denote the cell density, levoglucosan concentration, and ethanol concentration exhibited by *E. coli* LGE-*ycfR*, while Cell densities $_{CK}$, LG $_{CK}$, and EtOH $_{CK}$, denote those values exhibited by the non-overexpression strain *E. coli* LGE2.

concentrations of inhibitors like acetic acid, furfural, and phenol were also detected (Fig. 5E). Acetic acid concentration at the fermentation endpoint was slightly higher (0.09 g/L) than the original added concentration, possibly due to the acetic acid byproduct production during the fermentation. However, furfural was completely removed in the fermentation culture, and phenol concentration was slightly (0.07 g/L) lower than its original added concentration.

Finally, *E. coli* LGE-*ycfR* was subjected to the fermentation media supplied with combined inhibitors of the IC_{50} values identified for *E. coli* LGE2 (Fig. 5), namely 1.00 g/L acetic acid +0.20 g/L furfural +0.80 g/L phenol. The final cell density (OD_{600}), specific growth rate, ethanol concentration, ethanol productivity, and levoglucosan consumption rate of *E. coli* LGE-*ycfR* were improved by ~1.8, 2.5, 2.1, 2.6, and 2.3-folds of the control strain *E. coli* LGE2, respectively. The concentrations of the initially added inhibitors-acetic acid and phenol slightly increased and decreased by 0.13 and 0.05 g/L, respectively, while furfural was completely removed (Fig. 5E); these trends are similar to the above results presented by single inhibitor treated cells. Thus, our results show that overexpression of YcfR, which was revealed by our proteomics analysis to be a prominent stress-responsive protein, improved the inhibitor-resistance and fermentability of the resultant *E. coli* LGE-*ycfR* strain that was exposed to the biomass-derived inhibitors.

4. Discussion

The release of fermentable sugars from lignocellulosic biomass by pretreatment processes like pyrolysis is a crucial step in the biomass-based biochemicals and biofuels production industry. However, the pretreat procedures produce undesired byproducts like organic acids, furans, and phenols which are severely toxic and inhibitory to microbial growth and fermentation [24]. Especially, these inhibitors present in the biomass-derived liquid always exhibit synergistic effects on the fermenting strains rather than functioning independently [4]. Hence, the aims of this research were to reveal the proteins (genes) related to cellular resistance to multiple biomass-derived inhibitors and then develop a recombinant inhibitor-resistant strain to directly produce bioethanol from the levoglucosan-based and inhibitors-present media.

Consistent with the fact the cellular responses (i.e., the visible response is the change in cell growth at population level) to the same inhibitors are highly dependent on the strains and media used [4,25,26], our strain showed different IC_{50} and MIC values (Table 1). Therefore, determining the IC_{50} and MIC of individual and combined inhibitors is crucial for us to design the downstream sampling strategy for subsequent proteomics analysis. Further, our DIA-based quantitative proteomic analysis revealed the first insight into proteomics responses of the levoglucosan-utilizing and ethanol-producing *E. coli* strain to the representative biomass-derived inhibitors, identifying large amounts of DEPs involved in complex pathways and function categories (Figs. S1–S4 and Table S2, Supplementary file 1). We compared the DEPs presented by the four independent groups treated by individual and combined inhibitors, highlighted the universal DEPs shared by all the four groups, and consequently provided several inhibitors resistance related genes for further genetic manipulations. These universal DEPs including 15 up-regulated and 8 down-regulated proteins (Fig. 4A), were considered the biomarker proteins involved in cellular responses to various inhibitory stresses. *E. coli* itself induces inhibitor-detoxification processes via the regulation of stress-responsive and homeostasis mechanisms [27], therefore, regulating the expression of these stress-responsive proteins might benefit the biocatalysts to resist the inhibitors.

As a matter of fact, most of the identified biomarker proteins were the first time identified to be closely related to the biomass-derived inhibitors, although some of them have been reported to be involved in the responses to other different environmental stresses in a variety of strains. Among the 15 up-regulated biomarker proteins, outer membrane protein YcfR belonging to the YhcN family is a multiple stress resistance protein which has been proved to be up-regulated in response to heavy metals [28], low pH [29], heat shock [30], sodium salicylate [31], and H_2O_2 treatment [32]. CyoA, CyoB, and CyoC, as three cytochrome bo3 ubiquinol terminal oxidase subunits involved in the aerobic respiratory chain of *E. coli* for proton pumping and electron transfer, was reported to be up-regulated in response to low pH [29]. EvgS, as a sensor protein involved in the two-component signal transduction system in response to various environmental signals and stresses, has been proved to be up-regulated by a high concentration of alkali metals (Na^+, K^+) and low pH [33,34]. Besides, EvgS has also been identified as a regulator of the RND-type multidrug resistance efflux pump MdtF, which could respond to various toxic substances [35]. Ybl185, also named as ArcC (carbamate kinase), is involved in the arginine deiminase system which catalyzes the breakdown of arginine to ornithine and ammonia, with the production of ATP; ATP as an energy source and ammonium as a counteracter of the lowering pH resulting from sugar fermentation, could together contribute to the acid stress resistance of cells, even the cells membranes were already damaged [36,37]. PhoH is a phosphate starvation-inducible protein responding to the low nutrient condition [38] and has also been reported to be up-regulated in *Z. mobilis* in response to various biomass-derived inhibitors [4]. *E. coli* cells sense sugars and amino acids as attractants with counterclockwise flagellar motors rotation and sense weak acids like acetic acid as repellents with a clockwise pattern [39]. The bacterial chemotaxis related protein CheA, can transmit sensory signals from chemoreceptors to flagellar motors, causing swimming behavior in response to changes in cytoplasmic pH for cells to avoid acid crash [39]. Also, CheA that has been proved to be the convergence point of chemoreceptor methylation-dependent and -independent pathways for chemotactic adaptation, is also required for chemotaxis to oxygen, PTS (phosphotransferase system)-sugars, and non-PTS sugars [40]. Therefore, the up-regulation of CheA may also help *E. coli* resist the inhibitors by adjusting its cytoplasmic pH, energy production, and sugar uptake. Moreover, microbial cells need to import more nutrients into the cells and export toxins across membranes to protect the cells against inhibitory stresses. The ATP-binding cassette protein RbsA, as part of the ABC transporter complex RbsABC responsible for the transport of nutrients and pumping of toxins, has been proved to enhance the acid stress resistance of *Lactococcus lactis* when overexpressed in the cells [41]. KdsC involved in the biosynthesis of lipopolysaccharide and bacterial outer membrane synthesis [42], was reasonable to be up-regulated in response to the biomass-derived inhibitors, as it is evidenced by previous research that deletion of *kdsC* gene could lead to defective outer membrane and thus increase the cell sensitivity to hydrophobic toxic compounds [42]. YtfE has been shown to be responsible for the recovery and repair of nitrosylated Fe–S clusters damaged by oxidative and nitrosative stresses [43,44], thus the up-regulation of YtfE might also help *E. coli* to resist the inhibitors. High-affinity choline transport protein BetT, which was up-regulated, has also been proved to play a key role in response to the DNA damage stimulus [45] and hyperosmotic pressure [46]. In the 15 up-regulated DEPs, only the functional role of uncharacterized sugar kinase YgcE is not yet clear; however, our results suggest its role in cellular stress responses, in agreement with a previous report [47] that YgcE was upregulated by furfural stress.

In addition to the abovementioned proteins, the biomarker proteins like YebT, YfbS, GarL, AphA, NarJ, SbcB, MurR, and Ea22 were the first time to be identified as stress-responsive proteins. Interestingly, these proteins were all down-regulated responding to the inhibitors. YebT is an intermembrane transport protein related to the transport of substrates across the periplasm [48]; putative transporter YfbS with cation transmembrane transporter activity has a virulence-related function as a response regulator [49]; 2-Dehydro-3-deoxyglucarate aldolase GarL involved in D-galactarate and D-glycerate catabolism, is related to quorum sensing in *Pseudomonas savastanoi* and *Erwinia toletana* [50]; AphA, as a magnesium-dependent class B acid phosphatase, can catalyze the hydrolysis of various phosphoester substrates [51]. However, their

roles in stress-responses are also still unknown now. NarJ is involved in both membrane-bound respiratory nitrate reduction and nitrogen oxidation; it has been reported that interruption in the *narJ* gene inhibited the heterotrophic nitrification in *Pseudomonas* strain [52]. Thus, down-regulation of NarJ might contribute to *E. coli* by affecting the nitrification under the inhibitor stresses. Selective gene expression starts at the step of transcription initiation and replication [53]. SbcB mutation of *E. coli* could suppress the cell deficiency in postreplication repair by enabling the genes *uvrA*, *recB*, and *recC* to repair DNA double-strand breaks [54], suggesting that down-regulation of *sbcB* might be helpful for cells to repair the damaged DNA caused by inhibitory stresses. MurR is a putative transcription factor that could repress the *murPQ* operon, which is required for the cellular catabolism of the bacterial cell wall [55]. Also, Ea22, probably derived from bacteriophages, is an uncharacterized protein involved in the cell cycle control, cell division, and chromosome partitioning of *E. coli* [56]. Deletion-substitution mutagenesis has shown that Ea22 is responsible for the block to initiation of DNA replication [57]. Therefore, down-regulation of the transcription/replication-related proteins SbcB, MurR, and Ea22 might also help *E. coli* to overcome the adverse effects like DNA damage brought by the biomass-derived inhibitors via regulating the genetic process. Consequently, these new results provided a useful database of inhibitor-resistance related genes for the downstream metabolic engineering work and thus increased the knowledge depth in the context of inhibitor resistance of *E. coli*, although it has been studied a lot in recent years.

Based on the bioinformatics analysis above, the biomarker protein YcfR that exhibited the highest fold-change value under the challenge of various inhibitors, was chosen as the target for overexpression to further validate our proteomics results and develop a robust strain with high tolerance to the biomass-derived inhibitors. As expected, overexpression of gene *ycfR* coupled with heterologous expression of genes *lgk, pdc* and *adh* resulted in a recombinant inhibitor-tolerant strain *E. coli* LGE-*ycfR* that could successfully survive and produce ethanol from levoglucosan in the media provided with inhibitors at the MIC levels identified for the control strain *E. coli* LGE2. In minimal media, it has been previously shown that overexpression of acetyl-CoA synthetase (ACS) could make *E. coli* MG1655 (pNC5) grow on 64 mM (equal to 3.84 g/L) acetic acid [58]; random combination of deletion of gene *yqhD* and increased expression of *fucO*, *ucpA*, or *pntAB* could help ethanologenic *E. coli* LY180 tolerate 10–15 mM (equal to 0.96–1.44 g/L) furfural, and the maximal furfural tolerance (15 mM) was exhibited by strain XW129 [13,47,59,60]; and overexpression of whole phenol degradation pathway genes in *E. coli* BL221-AI can make the recombinant BL-*phe/cat* tolerate 10 mM (equal to 0.94 g/L) phenol and finally degrade 2 mM phenol [61]. In this regard, the engineered inhibitor-resistant strain *E. coli* LGE-*ycfR* was superior or comparable to the previous strains with one or multiple genetic changes. Furthermore, when challenged by single and combined inhibitors at their corresponding IC_{50} levels identified for the control strain, *E. coli* LGE-*ycfR* showed much lower sensitivity to the inhibitors, exhibiting higher final cell densities, specific growth rates, ethanol concentrations, ethanol productivities, and levoglucosan consumption rates than the control (Fig. 5). The ethanol yield reached a maximum of 0.42 g/g levoglucosan, equal to the yield without any added inhibitors [8]. Taken together, our results suggest that the overexpression of YcfR is an effective strategy for improving cellular tolerance to biomass-derived inhibitors.

It is well known that the biomass-derived inhibitors influence the cell growth and metabolism via causing the membrane, DNA, and protein damages, acid crash, and ROS accumulation, etc. By overexpressing gene *ycfR*, furfural added in the media was completely consumed (Fig. 5E) and transformed to furfuryl alcohol of less toxicity [13,47], implying that high concentration of furfural that cannot be detoxified in time could result in persistent DNA damage and NAD(P)H exhaustion in non-engineering strain [62] and overexpression of YcfR might have strengthened the regulation of this furfural-detoxification pathway, thereby decreasing the intracellular furfural pool. However, slight changes in the concentrations of acetic acid and phenol indicated that the acquirement of acetic acid and phenol tolerances of *E. coli* LGE-*ycfR* might be owing to other biological detoxification processes, i.e., membrane efflux pump [63] and amino acid dependent acid-resistance/tolerance system [64,65] rather than the direct transformation. YcfR is a small size protein (87 aa) that has previously been identified to be (1) a scaffolding or tethering protein [66], (2) a protein related to the cellular transport of metabolites [67], (3) part of the *soxRS* regulon protecting cells against superoxide and H_2O_2 stresses [67], and (4) a protein of high identity with the predicted stress-responsive cation transport ATPases of *Shigella boydii* and *E. coli* [67]. Previous researches have reported that YcfR can provide multiple stress resistances for cells to survive under various stress conditions like heavy metals [28], heat [30], sodium salicylate [31], and hydrogen peroxide treatment [32]. Deletion of gene *ycfR* in *E. coli* could render cells more hydrophobic and sensitive to environmental stresses [66], while increased level of YcfR could reduce the entry of toxic copper by reducing membrane permeability [66] and help cells to cope with the ambient stresses [67]. In addition, it is known that the homeostasis of lipids on the inner membrane and outer membrane plays crucial roles in drug resistance of Gram-negative bacteria [48]; interestingly, YcfR was also found to be a typical feature of outer membranes of Gram-negative organisms responsible for the transport of metabolites while not found in Gram-positive bacteria [66]. Also, due to its small size, it is inferred that YcfR might mainly play its roles in inhibitor tolerance mechanism by acting on other functional proteins (Fig. 6). As expected, most of the YcfR-interacting proteins, revealed by the bioinformatics data deposited in the BioGRID [20] database, were shown to help cells resist multiple stresses like acidic stress, oxidative stress, osmotic stress, and nutrient starvation stress, help cells repair the damaged membrane and DNA resulted from the toxic chemicals stimulus, and help cells degrade the inhibitors into less toxic compounds (Fig. 6). Therefore, YcfR could provide benefits to *E. coli* by playing roles in biomolecules transport, protein folding, DNA and membrane repair, and operon regulation, etc. Further study about overexpression or knockout of other biomarker proteins identified by our proteomics approach coupling with overexpression of *ycfR* might provide more benefits for the engineered *E. coli* to resist higher concentration of biomass-derived inhibitors.

5. Conclusions

Our comparative proteomics analysis revealed that 15 up- and 8 down-regulated proteins functioning in processes like biomolecules transport and metabolism, cell defense and signal transduction, and DNA repair were stress-responsive proteins candidate for the cellular tolerance to various inhibitors. Among all the candidate proteins, YcfR exhibiting the highest expression fold-change level was the most prominent protein in response to the biomass-derived inhibitors. Further overexpression of the gene *ycfR* coupled with heterologous expression of genes *lgk, pdc* and *adh* generated a recombinant *E. coli* LGE-*ycfR* strain, which exhibited better fermentation performances, cell growth profiles, and inhibitor-tolerance levels than the control strain. As a result, for the first time, we corroborated *ycfR* is a cellular resistance gene for multiple biomass-derived inhibitors, and developed an inhibitor-resistant levoglucosan-fermenting ethanologenic strain. The engineered strain that can eliminate the costly detoxicification processes commonly used to minimize the adverse effects of unavoidable inhibitory byproducts produced during the pretreatment of biomass, is promising for the *in situ* cost-effective bioethanol production from the thermo-decomposed biomass substrates.

CRediT authorship contribution statement

Dongdong Chang: Conceptualization, Methodology, Writing – original draft, Visualization. **Zia Ul Islam:** Data curation, Writing –

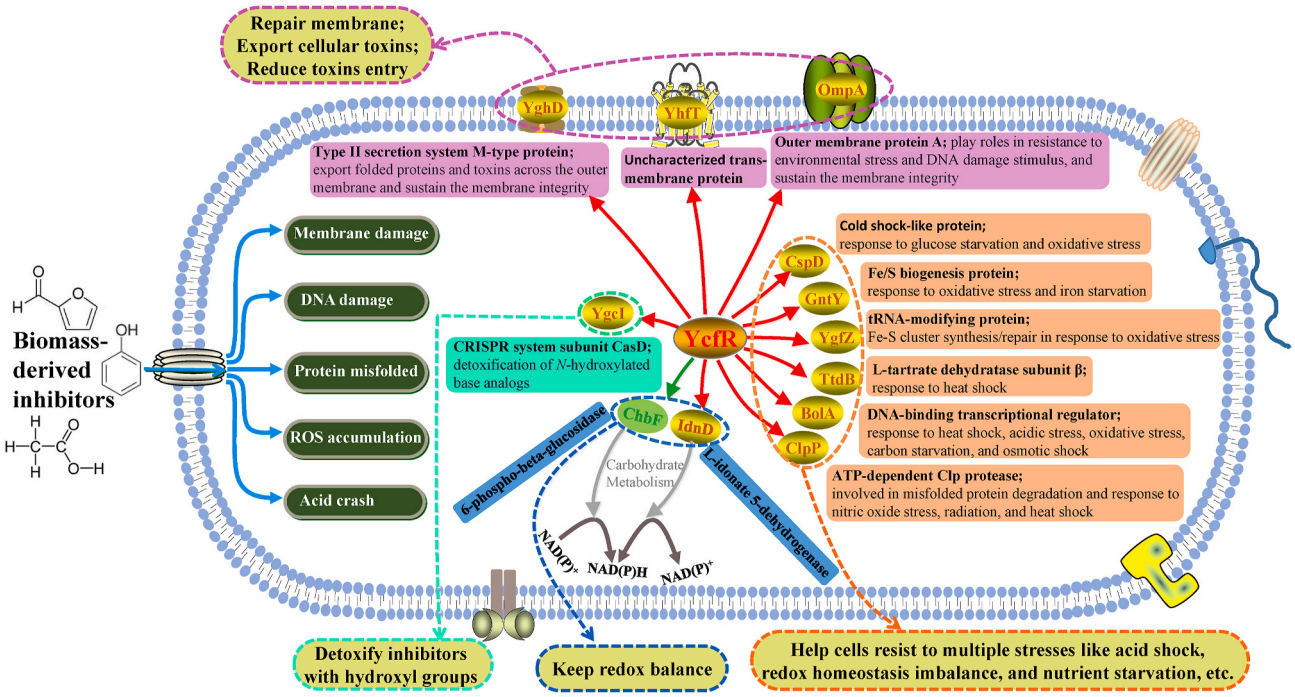


Fig. 6. Overview of the molecular functions of the YcfR-interacting proteins in *E. coli* against acetic acid, furfural, and phenol stresses. Red (green) arrows/nodes denote YcfR has negative (positive) interactions with the node protein, which mean that separate gene mutations/deletions combined in a same cell would result in a more (less) severe fitness defect or lethality than each gene did alone. Genetic/protein interactions were considered significant if they had an S-score≧3.08809 for positive interactions and S-score ≦ -3.38787 for negative interactions. The descriptions in the colored frames are the protein functions.

review & editing. **Junfang Zheng:** Software, Supervision. **Jie Zhao:** Investigation. **Xiaoyong Cui:** Supervision. **Zhisheng Yu:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work of this paper.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Natural Science Foundation of China (Grants NO. 21978287) and the Fundamental Research Funds for the Central Universities (Y954035XX2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2021.11.003.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its Supplementary files.

References

[1] Gupta V, Xue S, Jones AD, Sousa L, Piotrowski J, Jin M, Sarks C, Dale BE, Balan V. Water-soluble phenolic compounds produced from extractive ammonia pretreatment exerted binary inhibitory effects on yeast fermentation using synthetic hydrolysate. PLoS One 2018;13:e0194012.
[2] Rubin EM. Genomics of cellulosic biofuels. Nature 2008;454:841–5.
[3] Anex RP, Aden A, Kazi FK, Fortman J, Swanson RM, Wright MM, Satrio JA, Brown RC, Daugaard DE, Platon A. Techno-economic comparison of biomass-to-transportation fuels via pyrolysis, gasification, and biochemical pathways. Fuel 2010;89:S29–35.
[4] Chang D, Yu Z, Ul Islam Z, French WT, Zhang Y, Zhang H. Proteomic and metabolomic analysis of the cellular biomarkers related to inhibitors tolerance in *Zymomonas mobilis* ZM4. Biotechnol Biofuels 2018;11.
[5] Islam ZU, Zhisheng Y, Hassan el B, Dongdong C, Hongxun Z. Microbial conversion of pyrolytic products to biofuels: a novel and sustainable approach toward second-generation biofuels. J Ind Microbiol Biotechnol 2015;42:1557–79.
[6] Milne T, Agblevor F, Davis M, Deutch S, Johnson D. A review of the chemical composition of fast-pyrolysis oils from biomass Developments in thermochemical biomass conversion. Springer Netherlands; 1997. p. 409–24.
[7] Layton DS, Ajjarapu A, Choi DW, Jarboe LR. Engineering ethanologenic *Escherichia coli* for levoglucosan utilization. Bioresour Technol 2011;102:8318–22.
[8] Chang D, Islam ZU, Yang Z, Thompson IP, Yu Z. Conversion efficiency of bioethanol from levoglucosan was improved by the newly engineered Escherichia coli. Environ Prog Sustain Energy 2021:e13687.
[9] Chi Z, Rover M, Jun E, Deaton M, Johnston P, Brown RC, Wen Z, Jarboe LR. Overliming detoxification of pyrolytic sugar syrup for direct fermentation of levoglucosan to ethanol. Bioresour Technol 2013;150:220–7.
[10] Rover MR, Johnston PA, Jin T, Smith RG, Brown RC, Jarboe L. Production of clean pyrolytic sugars for fermentation. ChemSusChem 2014;7:1662–8.
[11] Bonomo J, Warnecke T, Hume PS, Marizcurrena A, Gill RT. A comparative study of metabolic engineering anti-metabolite tolerance in Escherichia coli. Metab Eng 2006;8:227–39.
[12] Alper HS, Stephanopoulos G. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 2007;9:258–67.
[13] Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO. Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 2013;110:4021–6.
[14] Li H, Zhang DF, Lin XM, Peng XX. Outer membrane proteomics of kanamycin-resistant *Escherichia coli* identified MipA as a novel antibiotic resistance-related protein. FEMS Microbiol Lett 2015:362.
[15] Clarkson SM, Hamiltonbrehm SD, Giannone RJ, Engle NL, Tschaplinski TJ, Hettich RL, Elkins J. A comparative multidimensional LC-MS proteomic analysis reveals mechanisms for furan aldehyde detoxification in *Thermoanaerobacter pseudethanolicus* 39E. Biotechnol Biofuels 2014;7:165. 165.
[16] Alonso-Gutierrez J, Kim EM, Batth TS, Cho N, Hu Q, Chan LJG, Petzold CJ, Hillson NJ, Adams PD, Keasling JD, Garcia Martin H, Lee TS. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. Metab Eng 2015;28:123–33.
[17] Luo H, Zheng P, Bilal M, Xie F, Zeng Q, Zhu C, Yang R, Wang Z. Efficient bio-butanol production from lignocellulosic waste by elucidating the mechanisms of *Clostridium acetobutylicum* response to phenolic inhibitors. Sci Total Environ 2020;710:136399.

[18] Narumi R, Masuda K, Tomonaga T, Adachi J, Ueda HR, Shimizu Y. Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics. Synth Syst Biotechnol 2018;3:97–104.

[19] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41:D808–15.

[20] Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47:D529–41.

[21] Chang D, Yu Z, Islam ZU, Zhang H. Mathematical modeling of the fermentation of acid-hydrolyzed pyrolytic sugars to ethanol by the engineered strain *Escherichia coli* ACCC 11177. Appl Microbiol Biotechnol 2015;99:4093–105.

[22] Bruschi M, KroMer JO, Steen JA, Nielsen LK. Production of the short peptide surfactant DAMP4 from glucose or sucrose in high cell density cultures of Escherichia coli BL21(DE3). Microb Cell Factories 2014;13:1–9.

[23] Jilani SB, Venigalla SSK, Mattam AJ, Dev C, Yazdani SS. Improvement in ethanol productivity of engineered *E. coli* strain SSY13 in defined medium via adaptive evolution. J Ind Microbiol Biotechnol 2017;44:1375–84.

[24] Islam ZU, Zhisheng Y, Hassan EB, Dongdong C, Hongxun Z. Microbial conversion of pyrolytic products to biofuels: a novel and sustainable approach toward second-generation biofuels. J Ind Microbiol Biotechnol 2015;42:1557–79.

[25] Mills TY, Sandoval NR, Gill RT. Cellulosic hydrolysate toxicity and tolerance mechanisms in Escherichia coli. Biotechnol Biofuels 2009;2:26. 26.

[26] Zaldivar J, Ingram LO. Effect of organic acids on the growth and fermentation of ethanologenic *Escherichia coli* LY01. Biotechnol Bioeng 2010;66:203–10.

[27] Park HS, Um Y, Sim SJ, Lee SY, Woo HM. Transcriptomic analysis of *Corynebacterium glutamicum* in the response to the toxicity of furfural present in lignocellulosic hydrolysates. Process Biochem 2015;50:347–56.

[28] Egler M, Grosse C, Grass G, Nies DH. Role of the extracytoplasmic function protein family sigma factor RpoE in metal resistance of Escherichia coli. J Bacteriol 2005;187:2297.

[29] Maurer LM, Yohannes E, Bondurant SS, Radmacher M, Slonczewski JL. pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. J Bacteriol 2005;187:304–19.

[30] Richmond C, Glasner JD, Mau R, Jin H, Blattner FR. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res 1999;27:3821–35.

[31] Pomposiello PJ, Bennik MHJ, Demple B. Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. J Bacteriol 2001;183:3890–902.

[32] Zheng M, Wang X, Templeton LJ, Smulski DR, Larossa RA, Storz G. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. J Bacteriol 2001;183:4562–70.

[33] Eguchi Y, Utsumi R. Alkali metals in addition to acidic pH activate the EvgS histidine kinase sensor in Escherichia coli. J Bacteriol 2014;196:3140–9.

[34] Sen H, Aggarwal N, Ishionwu C, Hussain N, Parmar C, Jamshad M, Bavro VN, Lund PA. Structural and functional analysis of the *Escherichia coli* acid-sensing histidine kinase EvgS. J Bacteriol 2017:199.

[35] Nishino K, Yamaguchi A. EvgA of the two-component signal transduction system modulates production of the YhiUV multidrug transporter in Escherichia coli. J Bacteriol 2002;184:2319–23.

[36] Casianocolon A, Marquis RE. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. Appl Environ Microbiol 1988;54:1318–24.

[37] Kok J, Buist G, Zomer A, Van Hijum SAFT, Kuipers OP. Comparative and functional genomics of lactococci. FEMS Microbiol Rev 2005;29:411–33.

[38] Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, Carlson CA, Smith RD, Giovanonni SJ. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J 2009;3:93–105.

[39] Conley MP, Berg HC, Tawa P, Stewart RC, Ellefson DD, Wolfe AJ. pH dependence of CheA autophosphorylation in Escherichia coli. J Bacteriol 1994;176:3870–7.

[40] Rowsell EH, Smith JM, Wolfe A, Taylor BL. CheA, CheW, and CheY are required for chemotaxis to oxygen and sugars of the phosphotransferase system in Escherichia coli. J Bacteriol 1995;177:6011–4.

[41] Zhu Z, Yang J, Yang P, Wu Z, Zhang J, Du G. Enhanced acid-stress tolerance in *Lactococcus lactis* NZ9000 by overexpression of ABC transporters. Microb Cell Factories 2019;18:136.

[42] Sperandeo P, Pozzi C, Deho G, Polissi A. Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the *Escherichia coli yrbG-yhbG* locus. Res Microbiol 2006;157:547–58.

[43] Silva LSO, Baptista JM, Batley C, Andrews SC, Saraiva LM. The Di-iron RIC protein (YtfE) of *Escherichia coli* interacts with the DNA-binding protein from starved cells (dps) to diminish RIC protein-mediated redox stress. J Bacteriol 2018;200.

[44] Justino MC, Almeida CC, Teixeira M, Saraiva LM. *Escherichia coli* Di-iron YtfE protein is necessary for the repair of stress-damaged iron-sulfur clusters. J Biol Chem 2007;282:10352–9.

[45] Khil PP, Cameriniotero RD. Over 1000 genes are involved in the DNA damage response of Escherichia coli. Mol Microbiol 2002;44:89–105.

[46] Tondervik A, Strom AR. Membrane topology and mutational analysis of the osmotically activated BetT choline transporter of Escherichia coli. Microbiology 2007;153:803–13.

[47] Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, York SW, Nunn D, Shanmugam KT, Ingram LO. Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol 2009;75:6132–41.

[48] Liu C, Ma J, Wang J, Wang H, Zhang L. Cryo-EM structure of a bacterial lipid transporter YebT. J Mol Biol 2020;432:1008–19.

[49] Chattopadhyay S, Paul S, Kisiela DI, Linardopoulou EV, Sokurenko EV. Convergent molecular evolution of genomic cores in *Salmonella enterica* and *Escherichia coli*. J Bacteriol 2012;194:5002.

[50] Caballo-Ponce E, Meng X, Uzelac G, Halliday N, Venturi V. Quorum sensing in *Pseudomonas savastanoi* pv. savastanoi and *Erwinia toletana*: role in virulence and interspecies interactions in the olive knot. Appl Environ Microbiol 2018;84. AEM.00950-18.

[51] Leone R, Cappelletti E, Benvenuti M, Lentini G, Thaller MC, Mangani S. Structural insights into the catalytic mechanism of the bacterial class B phosphatase AphA belonging to the DDDD superfamily of phosphohydrolases. J Mol Biol 2008;384:478–88.

[52] Nemergut DR, Schmidt SK. Disruption of *narH, narJ*, and *moaE* inhibits heterotrophic nitrification in *Pseudomonas* strain M19. Appl Environ Microbiol 2002;68:6462–5.

[53] Browning DF, Busby SJ. Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol 2016;14:638–50.

[54] Wang TV, Smith KC. Mechanism of *sbcB*-suppression of the *recBC*-deficiency in postreplication repair in UV-irradiated *Escherichia coli* K-12. Mol Genet Genom 1985;201:186–91.

[55] Ishii D, Ishihama A, Yamamoto K. Two modes of autoregulation of the *murR* repressor in *Escherichia coli*. Biosci Biotechnol Biochem 2009;73:2528–30.

[56] Tong J, Nejman-Faleńczyk B, Bloch S, Węgrzyn A, Donaldson LW. Ea22 proteins from lambda and shiga toxin-producing bacteriophages balance structural diversity with functional similarity. ACS Omega 2020;5:12236–44.

[57] Sergueev K, Court D, Reaves L, Austin SE. *Coli* cell-cycle regulation by bacteriophage lambda. J Mol Biol 2002;324:297–307.

[58] Lin H, Castro NM, Bennett GN, San KY. Acetyl-CoA synthetase overexpression in *Escherichia coli* demonstrates more efficient acetate assimilation and lower acetate accumulation: a potential tool in metabolic engineering. Appl Microbiol Biotechnol 2006;71:870–4.

[59] Wang X, Miller EN, Yomano LP, Zhang X, Shanmugam KT, Ingram LO. Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. Appl Environ Microbiol 2011;77:5132–40.

[60] Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic Escherichia coli. Appl Environ Microbiol 2009;75:4315.

[61] Wang B, Xu J, Gao J, Fu X, Han H, Li Z, Wang L, Tian Y, Peng R, Yao Q. Construction of an *Escherichia coli* strain to degrade phenol completely with two modified metabolic modules. J Hazard Mater 2019;373:29–38.

[62] Mills TY, Sandoval NR, Gill RT. Cellulosic hydrolysate toxicity and tolerance mechanisms in Escherichia coli. Biotechnol Biofuels 2009;2:26.

[63] Zhang DF, Li H, Lin XM, Wang SY, Peng XX. Characterization of outer membrane proteins of *Escherichia coli* in response to phenol stress. Curr Microbiol 2011;62:777–83.

[64] Foster JW. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat Rev Microbiol 2004;2:898–907.

[65] Xu Y, Zhao Z, Tong W, Ding Y, Liu B, Shi Y, Wang J, Sun S, Liu M, Wang Y, Qi Q, Xian M, Zhao G. An acid-tolerance response system protecting exponentially growing Escherichia coli. Nat Commun 2020;11:1496.

[66] Mermod M, Magnani D, Solioz M, Stoyanov JV. The copper-inducible ComR (YcfQ) repressor regulates expression of ComC (YcfR), which affects copper permeability of the outer membrane of Escherichia coli. Biometals 2012;25:33–43.

[67] Zhang XS, Garcia-Contreras R, Wood TK. YcfR (BhsA) influences *Escherichia coli* biofilm formation through stress response and surface hydrophobicity. J Bacteriol 2007;189:3051–62.