D M M S -F
Zymomonas mobilis R P

L - C , -F , -H C , J- , J ,
C -G L , S -H , R F - B

aState Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai, China

bJoint International Research Laboratory of Metabolic & Developmental Science of the Ministry of Education, Shanghai Jiao Tong University, Shanghai, China

cSchool of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

dState Key Laboratory of Biocatalysis and Enzyme Engineering, Hubei University, Wuhan, China

eSchool of Life Science, Hubei University, Wuhan, China

fKarolinska Institutet, Stockholm, Sweden

ABSTRACT *Zymomonas mobilis* metabolizes sugar anaerobically through the Entner-Doudoroff pathway with less ATP generated for lower biomass accumulation to direct more sugar for product formation with improved yield, making it a suitable host to be engineered as microbial cell factories for producing bulk commodities with major costs from feedstock consumption. Self-flocculation of the bacterial cells presents many advantages, such as enhanced tolerance to environmental stresses, a prerequisite for achieving high product titers by using concentrated substrates. ZM401, a self-flocculating mutant developed from ZM4, the unicellular model strain of *Z. mobilis*, was employed in this work to explore the molecular mechanism underlying this self-flocculating phenotype. Comparative studies between ZM401 and ZM4 indicate that a frameshift caused by a single nucleotide deletion in the poly-T tract of ZMO1082 fused the putative gene with the open reading frame of ZMO1083, encoding the catalytic subunit BcsA of the bacterial cellulose synthase to catalyze cellulose biosynthesis. Furthermore, the single nucleotide polymorphism mutation in the open reading frame of ZMO1055, encoding a bifunce.6(ITJ/F712(e)-3(a)-.6(n)-417.9t-6.1(.8(e-.0474.9(3)]TJ/(2.3(tiF2.5(3.1(r)2.E.1(.8A17.9t-L)-2.5(4(a).52.1(s)2.17(g)teth)3.7(e)-506(i))-44)]TJT-3

ZMac39.9(B)-2.9(d-4.1,f)-1(-)]l.5197ead(y)1ntee.5(i)4(sow)-80t

In addition to unicellular cells, bacteria can form biofilms, in particular, for pathogens to resist antibacterial agents and the immune system through social cooperation that can be mediated by quorum sensing (QS) signal molecules (1–3). The matrix of biofilms is composed of extracellular polymeric substances (EPS), including polysaccharides, nucleic acids, proteins, and lipids produced and secreted by the bacterial cells, and a developmental life cycle from formation to maturation to dispersion is generally observed for biofilms (4, 5). Activated sludge, which has been employed with a long history in wastewater treatment, is another morphology formed naturally with abundant bacteria as core communities through symbiosis, EPS, and other materials (6–9). Under pure culture conditions, some bacterial cells can self-flocculate to form multicellular flocs that are significantly different in morphology from amorphous biofilms developed generally on surfaces, as well as activated sludge formed by different species with spatial heterogeneity.

Bacterial flocs present many advantages for industrial applications. On the one hand, bacterial flocs can settle down quickly through gravity sedimentation for cost-effective biomass recovery (see Video S1 in the supplemental material). On the other hand, when self-flocculating, bacterial cells can be immobilized within bioreactors for high cell density to improve productivities as highlighted in ethanol fermentation using self-flocculating yeast (10). Most importantly, bacterial flocs are more tolerant to stresses, a prerequisite for robust production, particularly for producing biofuels, bioenergy, and bio-based chemicals from lignocellulosic biomass with concentrated substrates for high product titers, since various inhibitors are released during the pretreatment of the biomass (11).

Compared to the Embden-Meyerhof-Parnas (EMP) pathway, which is commonly employed by other species, *Zymomonas mobilis* metabolizes sugar anaerobically through the Entner-Doudoroff (ED) pathway with 50% less ATP generated for lower biomass accumulation so that more sugar can be directed to product formation with high yield (12). Meanwhile, the low energy-coupling respiration in *Z. mobilis* accelerates metabolic flux with the ED pathway, which consequently facilitates its production rate under low-biomass-accumulation conditions (13). As a result, *Z. mobilis* has been acknowledged as a chassis to be engineered as microbial cell factories for biorefineries to produce bulk commodities such as biofuels and bio-based chemicals with major costs from feedstock consumption (14, 15).

As an ethanologenic species, *Z. mobilis* is more suitable for cellulosic ethanol production (16), which was pioneered by DuPont in Iowa, USA, since it can be engineered to metabolize pentose sugars released during the hydrolysis of hemicelluloses in lignocellulosic biomass through the isomerase pathway without cofactor imbalance (17, 18), an intrinsic disadvantage in engineering *Saccharomyces cerevisiae* for the same purpose, which results in intermediate accumulation to compromise ethanol yield (19). However, microbial cells are suffering from various stresses under production conditions, and studies have been performed on engineering *Z. mobilis* with tolerance to accumulated ethanol, acetic acid released during the pretreatment of lignocellulosic biomass, and high temperature, which is preferred for industrial fermentation (20–22), but unfortunately, the progress is less significant, since multiple stresses are always presented simultaneously during industrial production.

The self-flocculating mutant ZM401 developed from ZM4, the unicellular model strain of *Z. mobilis*, showed improved tolerance to inhibitors released during the pretreatment of lignocellulosic biomass (23). Such a feature is extremely important for producing cellulosic ethanol and bio-based chemicals directly from the hydrolysate of lignocellulosic biomass without a necessity for detoxification, which is too costly for producing bulk commodities with marginal revenues (24, 25). Our previous studies indicated that the basis for the self-flocculation of ZM401 is the production of cellulose fibrils as extracellular matrix components (26), but the molecular mechanism underlying the self-flocculation of the bacterial cells is still unclear.

Bioinformatic analysis indicated that the genome of ZM4 contains a bacterial cellulose synthase (*bcs*) operon (27), but the bacterial cells seem not to be able to synthesize cellulose fibrils for self-flocculation. We therefore hypothesized that mutations in

the *bcs* operon encoding enzymes and structural components to catalyze cellulose biosynthesis and/or other genes encoding enzymes to regulate the biosynthesis could occur in the genome of ZM401, making the mutant developed with the self-flocculating phenotype. In this work, we performed comparative genome and transcriptome analyses between ZM401 and ZM4 to explore the underlying molecular mechanism for the self-flocculation. The analyses and experimental validations indicated that several alterations in the genome are required for the self-flocculation of ZM401. The progress is significant for engineering unicellular strains from *Z. mobilis* with the self-flocculating phenotype through rational design for robust production.

RESULTS

Genetic basis for the self-flocculating phenotype of ZM401. Combining the data collected from long-read and short-read sequencing, we assembled the genome of ZM401 to explore genetic basis for the self-flocculation of the bacterial cells through comparative genomic analysis with the genome of ZM4 as the reference (28–30). The genome of ZM401 contains one circular chromosome of 2,058,754 bp and four circular plasmids of 32,791 bp (401_pZM32), 33,006 bp (401_pZM33), 36,494 bp (401_pZM36), and 39,266 bp (401_pZM39).

One nucleotide deletion was detected in the circular chromosome of ZM401, but the size of four circular plasmids is identical to those in ZM4. We identified a total of 35 single nucleotide variations (SNVs) in ZM401, but no rearrangement was observed in the genome. For the 35 SNVs, the nucleotide deletion occurred in the putative gene ZMO1082, and 34 single nucleotide polymorphisms (SNPs) occurred in other genes and intergenic regions (Table S1). Verification of the SNVs by Sanger sequencing confirmed that the SNP within ZMOp32×018 in the plasmid ZMOp32 at the site of 1789 and another SNP within the intergenic region at the site of 1614973 in the circular chromosome in ZM401 corresponded to nucleotides at the site of 1789 and 1614974 in the genome of ZM4, indicating that either the genome sequence of our ZM4 strain from ATCC might mutate at these two sites or the database information of ZM4 needs to be corrected. We therefore confirmed, in total, 33 SNVs in the genome of ZM401 compared to ZM4, 31 of which are in the genes and 2 of which are located in untranslated regions. For the 31 SNVs in the genes, one is a nucleotide deletion in ZMO1082, and all others are SNPs, which include 17 nonsynonymous mutations (Table S2) and 13 synonymous mutations.

In order to identify genes involved in controlling the self-flocculating phenotype of ZM401, we performed comparative transcriptomic analysis between ZM401 and ZM4. The transcriptome analysis through RNA sequencing (RNA-seq) indicated that, in total, 74 genes were differentially expressed in ZM401, whereby 63 genes were upregulated, and 11 genes were downregulated (Table S3). Among the 18 genes with nonsilent mutations in ZM401, only ZMO1082 containing the nucleotide deletion was significantly upregulated, and no differential expression occurred for all the 17 genes with nonsynonymous SNPs. Subsequently, we performed gene ontology enrichment for those differentially expressed genes to further explore their functions (Fig. S1).

For those upregulated genes, the largest set is with the cell components, in particular, the cell membrane, followed by the set with biological processes associated with metabolism for those major cell wall components and the set with molecular functions such as transmembrane activities. In particular, genes conducting the biosynthesis of two major β-glycans, 1,4-β-glucan cellulose and levan, a secreted exopolysaccharide consisting of 2,6-β-linked fructose subunits, are upregulated (Table S3). These results support our hypothesis for the self-flocculation of ZM401—the direct interactions of EPS synthesized predominantly under the catalysis of membrane-embedded enzymes such as cellulose synthase for cellulose biosynthesis (26). On the other hand, the functions of the 11 downregulated genes are predominately related to development and metabolism, including DNA biosynthesis and replication as well as oxidoreductase activities, which are significantly different from the functions observed for the upregulated genes, indicating their smaller impact on the self-

flocculation of ZM401. We further predicted protein-protein interactions for those significantly upregulated genes using the software STRING and DOOR[2] (31, 32), and ZMO1082, ZMO1083, ZMO1084, and ZMO1085 were detected as potential interaction partners (Fig. S2 and S3).

Although many genes in the genome of *Z. mobilis* are still putative, previous analyses indicated that ZM4 contains a *bcs* gene cluster with ZMO1083, ZMO1084, ZMO1085, and ZMO1086, encoding a membrane-integrated bacterial cellulose synthase complex composed of subunits BcsA and BcsB, the outer membrane channel BcsC, and the cellulase BcsZ, respectively. The catalytic subunit BcsA for cellulose biosynthesis is frequently fused with BcsB, a large periplasmic protein embedded with its C terminus in the inner membrane guiding the export of synthesized cellulose across the periplasm toward the outer membrane (27, 33). Missense mutations in BcsA can lead to changes in the biosynthesis of cellulose (34).

ZMO1082 has been predicted to be a putative gene in ZM4, encoding a peptide of 67 amino acids with no predicted functionality on its own (29). The single nucleotide deletion from the homopolymeric tract of nine thymines in ZM401 coincidently affected ZMO1083, since the base deletion resulted in a frameshift mutation, which disrupted the stop codon TGA of ZMO1082, and produced a fusion with ZMO1083 for a new gene, ZMO1083/2, to encode BcsA in ZM401 (Fig. 1a). We predict that, with 60 additional amino acids at the N terminus, the catalytic function of BcsA encoded by the new gene has been enhanced for more efficient synthesis of cellulose, which further develops as cellulose fibrils by the Bcs complex of ZM401 to flocculate the bacterial cells (Fig. 1b).

Cyclic diguanylic acid (c-di-GMP) is a second messenger for signal transduction in bacteria which regulates intracellular processes, including cellulose biosynthesis (35, 36). Bacteria synthesize c-di-GMP from GTP catalyzed by diguanylate cyclases with GGDEF domains, but phosphodiesterases (PDE) with EAL domains can degrade c-di-GMP (37). ZMO1055 in ZM4 is predicted to encode the bifunctional enzyme diguanylate cyclase/phosphodiesterase with the conserved/semiconserved motifs of GGDQF and EAL, respectively, to catalyze the synthesis and degradation of c-di-GMP. Although the transcription of ZMO1055 did not change
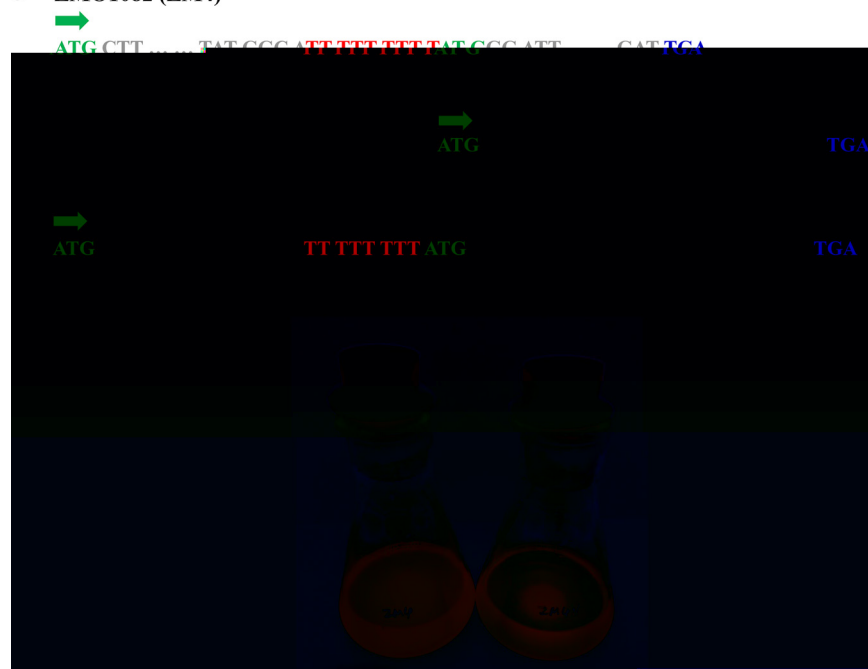
a **ZMO1082 (ZM4)**

FIG 1 Nucleotide sequences for the putative gene ZMO1082 and the gene ZMO1083 in ZM4 and their integration for the new gene ZMO1083/2 in ZM401 (a) and the morphologies of the unicellular model strain ZM4 and the self-flocculating mutant ZM401 (b). The dynamic gravity sedimentation of the bacterial flocs is highlighted in Video S1.

significantly in ZM401, the amino acid substitution created by the SNP mutation is located in the EAL domain, which could compromise the catalytic activity of the phosphodiesterase. Thus, c-di-GMP degradation would be impaired for its accumulation to regulate intracellular processes more effectively, in particular, cellulose biosynthesis for the self-flocculation of the bacterial cells. The upregulation of all genes with the *bcs* gene cluster, with the exception of ZMO1086 only in ZM401, partly supported this speculation.

Molecular docking for interactions of enzymes with substrates and activators. We were interested to know whether the cellulose synthase BcsA and the GGDEF-EAL protein ZMO1055 in ZM401 could show features that would point to altered catalytic activities.

For bacterial cellulose synthesis, the active site of BcsA is stimulated first through the repositioning of its gating loop by c-di-GMP, and then the substrate UDP-glucose binds onto the activated site for glucose to be cleaved and transferred to extend the growing polysaccharide chain at its nonreducing end with UDP released. Subsequently, UDP needs to be replaced by UDP-glucose for the biosynthesis to continue, but the repositioning of the gating loop to either open the catalytic pocket for accommodating UDP-glucose or coordinating c-di-GMP at the active site is UDP-dependent (38). We therefore simulated interactions between BcsA encoded by ZMO1083 in ZM4 or ZMO1083/2 in ZM401 with UDP to assess their impact on cellulose biosynthesis. The simulation results indicate that when docking with UDP, ZMO1083 in ZM4 and ZMO1083/2 in ZM401 exhibited similar affinity, which was quantified by their LibDock scores of 121.8 and 121.7, respectively, but when docking with c-di-GMP, the LibDock score for ZMO1083/2 is 124.7 compared to that of 100.1 only for ZMO1083, suggesting that ZMO1083/2 would dock with c-di-GMP effectively to activate cellulose biosynthesis more efficiently (Fig. 2).

The EAL domain is conserved in Gram-negative bacteria, including *Z. mobilis*, for catalyzing c-di-GMP degradation (35). The amino acid substitution Ala526Val in ZMO1055 with ZM401 is located in the C terminus of the EAL domain with the phosphodiesterase, which could affect its binding with c-di-GMP. To test this hypothesis, we performed structural modeling for proteins encoded by ZMO1055 in ZM401 and ZM4 (Fig. 3a). The LibDock score for ZMO1055 in ZM401 and ZM4 is 109.4 and 103.3, respectively, indicating that the single amino acid mutation could not change the protein structure significantly to affect its affinity for c-di-GMP. We therefore calculated binding stability for c-di-GMP with the EAL domain, which depends on various noncovalent interactions such as
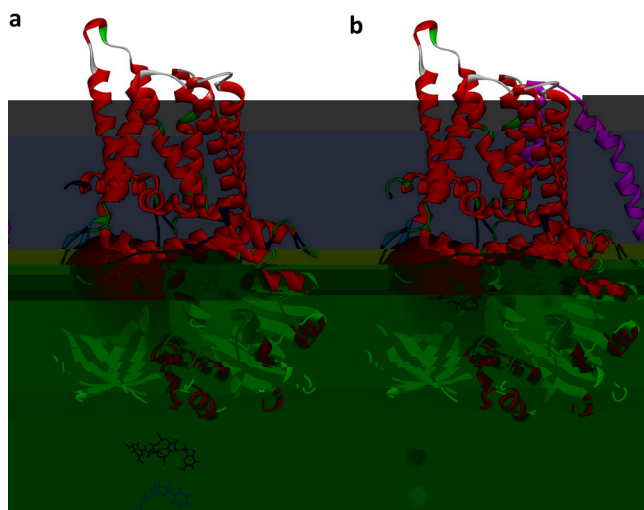

FIG 2 Molecular docking for UDP and c-di-GMP with BcsA encoded by ZMO1083 in ZM4 (a) and ZMO1083/2 in ZM401 (b). Bright purple, mutation in the domain; light blue and orange balls, regions with low free energy for recruiting UDP-glucose with UDP as a product released and c-di-GMP as the signal molecule to mediate cellulose biosynthesis, respectively.
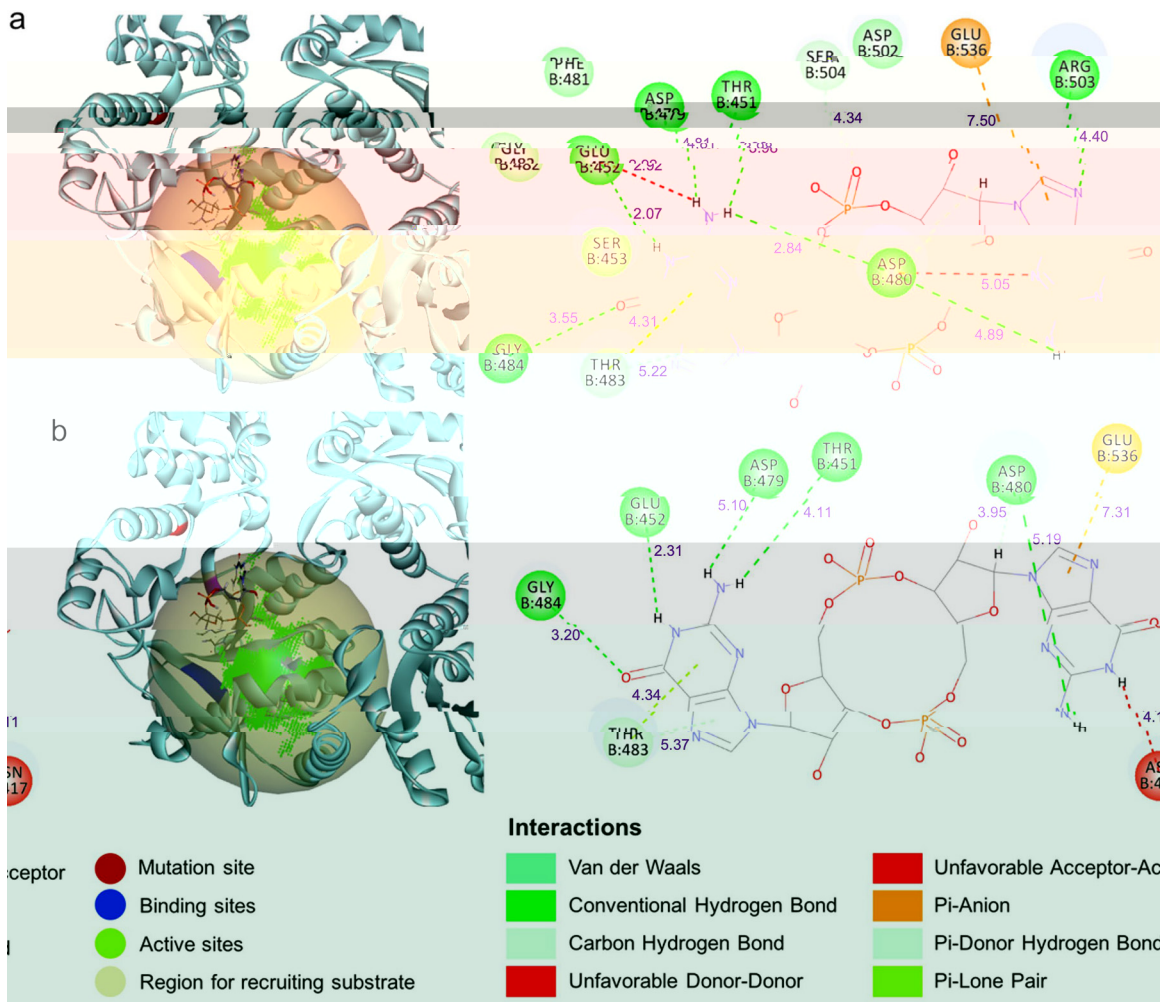
FIG 3 Molecular docking for c-di-GMP with the active center of the EAL domain in the bifunctional enzyme encoded by ZMO1055 in ZM4 (a) and ZM401 (b). The numbers are distances in angstrom (Å).

Van der Waals forces, π bonds formed with negatively charged GLU536 (pi-anion), hydrogen bonds developed with THR483 through π bonds as electron donors (pi-donor hydrogen bond), and conventional hydrogen bonds with ASP480 and ARG503 (Fig. 3b). Although the mutation Ala526Val could not affect most of these interactions, it seems to affect the formation of hydrogen bonds between c-di-GMP and the two amino acid residues ASP480 and ARG503 of the EAL domain.

For ZMO1055 in ZM4, ASP480 could cyclize with c-di-GMP through two hydrogen bonds to stabilize the docking, making the EAL domain effective for its degradation, but the docking of c-di-GMP into the EAL domain in ZM401 seems less stable, since ASP480 would link with c-di-GMP through one hydrogen bond only. On the other hand, another hydrogen bond that might develop between c-di-GMP and ARG503 in ZM4 could not form in ZM401. We therefore calculated binding energy between ZMO1055 and c-di-GMP, and the results indicate that very low energy of −0.01 kcal/mol is needed for the wild-type ZMO1055 to bind with c-di-GMP, but the binding energy increases to 8.12 kcal/mol when the mutation occurs (Fig. S4). As a result, the degradation of c-di-GMP could be compromised in ZM401, resulting in its intracellular accumulation to activate the synthesis of cellulose fibrils to flocculate the bacterial cells (38).

Experimental validations. When either ZMO1082 or ZMO1083 of ZM401 was deleted, the chord length measured by the focused beam reflectance measurement (FBRM) system decreased from 425.6 μm detected for the bacterial flocs to an undetectable level for unicellular bacterial cells, indicating that the self-flocculating phenotype of ZM401 was disrupted

completely; however, when ZMO1083/2 was overexpressed in ZM4, no self-flocculation of
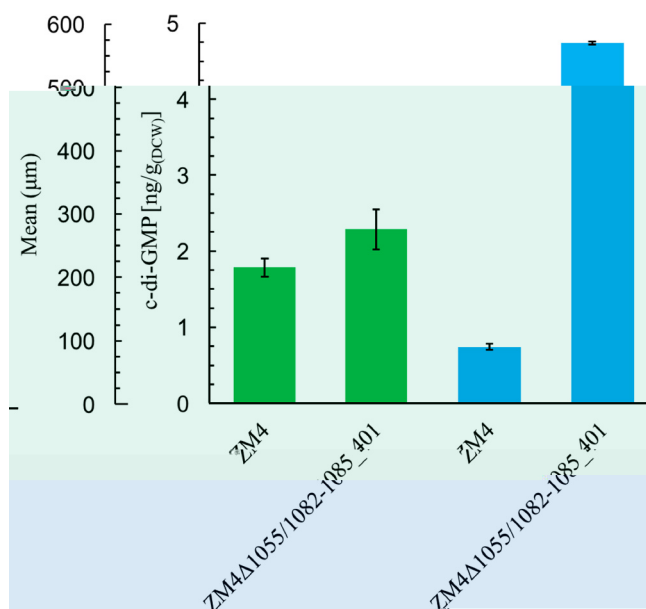
FIG 5 Engineering ZM4 with the self-flocculating phenotype through the knockout of ZMO1055 and the overexpression of the *bcs* operon from ZM401 (green, c-di-GMP; and blue, mean). The chord length distribution of the bacterial flocs and their dynamic gravity sedimentation are highlighted in Fig. S7 and Video S2.

increased to 3.88 ng/g (DCW) compared to that of 1.78 ng/g (DCW) detected in the wild-type strain, which suggested that the phosphodiesterase catalytic activity of the EAL domain of ZMO1055 significantly impacts the degradation of c-di-GMP. The intracellular c-di-GMP of 2.87 ng/g (DCW) detected in ZM401 was significantly higher than that in the wild-type ZM4, but lower than that in the ZMO1055 deletion mutant of ZM4. When ZM401 was engineered with the overexpression of ZMO1055 from ZM4 driven by the strong promoter, the intracellular accumulation of c-di-GMP decreased drastically to 0.17 ng/g (DCW). However, no significant change in the intracellular accumulation of c-di-GMP was observed when its own ZMO1055 was overexpressed. This finding suggests that the Ala526Val substitution in ZMO1055 compromised its PDE activity, and thus a much higher expression level is needed to degrade c-di-GMP significantly. These experimental results clearly validated that the SNP mutation with the open reading frame of ZMO1055 in the genome of ZM401, leading to the Ala526Val substitution, substantially compromised the catalytic function of the EAL domain for c-di-GMP degradation. The intracellular accumulation of c-di-GMP could consequently enhance the biosynthesis of cellulose fibrils by the Bcs complex with BcsA created by the fusion of ZMO1083 with ZMO1082 encoded by ZMO1083/2 to induce the self-flocculation of the bacterial cells.

Engineering ZM4 with the self-flocculating phenotype. With the roles of ZMO1083/2 and ZMO1055 in the self-flocculation of ZM401 validated experimentally, we deleted ZMO1055 from ZM4 to enhance its intracellular accumulation of c-di-GMP, followed by the overexpression of the *bcs* operon from ZM401 for the biosynthesis of cellulose fibrils.

The engineered strain ZM4Δ1055/BcsABC_401 developed the self-flocculating phenotype, which was characterized by a chord length of 568.5 μm for the bacterial flocs and an intracellular accumulation of c-di-GMP up to 2.28 ng/g (DCW), which is significantly higher than that of the 1.78 ng/g (DCW) detected in the wild-type ZM4, to activate cellulose biosynthesis catalyzed by BcsABC_401 for the self-flocculation of the bacterial cells (Fig. 5). The dynamic gravity sedimentation of ZM4Δ1055/BcsABC_401 is highlighted by the video (Video S2). These results also indicate that under these experimental conditions, BcsZ is not required for the self-flocculation of the ZMO1055 deletion mutant of ZM4 upon overexpression of the BcsABC operon of ZM401.

DISCUSSION

Since the genome of ZM4 was sequenced and reported in 2005 (28), many studies have been performed on metabolic engineering of *Z. mobilis* to exploit its merits associated with the ED pathway for glycolysis, through which less ATP is produced for low biomass accumulation to improve product yield (15, 16), in particular, for producing bulk commodities such as ethanol as a biofuel and 2,3-butanediol, a bio-based chemical as feedstock to produce synthetic rubber, plasticizer, and other products (39, 40).

Like other bacteria such as *Komagataeibacter xylinus* and *Escherichia coli* (41), the genome of *Z. mobilis* contains an identifiable variant of the *bcs* operon (27), but all strains from this species except ZM401 are unicellular. Some bacteria can develop biofilms, and surfaces are needed in general for bacterial cells to attach onto to form biofilms (2). While the EPS matrix of biofilms can frequently contain cellulose, additional components such as proteins, lipids, and nucleic acids are also present to glue bacterial cells together for the amorphous architecture (5). In contrast, to the best of our knowledge, the only EPS required for the self-flocculation of *Z. mobilis* is cellulose.

The self-flocculating phenotype of bacterial cells is controlled ultimately at the molecular level, but a dynamic balance can develop between the self-flocculation of the bacterial cells and the breakup of the bacterial flocs by shearing from mixing, such as hydrodynamic shearing in shaking flasks and both hydrodynamic and mechanical shearing created through agitation within bioreactors. The larger the bacterial flocs are, the poorer their resistance to shearing force will be for being broken up into small ones to better resist the shearing force. On the other hand, those smaller bacterial flocs, when contacted, can flocculate together again to form large ones. Not only can such a dynamic process generate a broad size distribution for the bacterial flocs, but their insides can also be renewed timely for sustained viability (Fig. S8), making them more suitable for industrial production, particularly for continuous culture and fermentation to produce biofuels and other bulk products in large quantities, which can be operated for a long time, even year-long without interruption.

Transcriptomic analysis suggested that the *bcs* operon required for cellulose biosynthesis is responsible for the self-flocculation of ZM401 (27), and experimental studies validated such a speculation (26), raising the question: why the *bcs* operon in ZM4 cannot synthesize cellulose to flocculate the bacterial cells like ZM401. Comparative genomic and transcriptomic analyses and molecular docking simulations for interactions of enzymes with substrates and activators indicate that the frameshift mutation created by the nucleotide deletion from a polynucleotide tract in the open reading frame of ZMO1082 created a novel BcsA fusion protein with an extended N terminus in ZM401, which shows improved catalytic activities on cellulose biosynthesis for the mutant to develop the self-flocculating phenotype. While ZMO1083/2 formed through the fusion of ZMO1082 into ZMO1083 is the prerequisite for synthesizing cellulose fibrils to flocculate the bacterial cells, the SNP mutation on ZMO1055 that compromises c-di-GMP degradation is suggested to contribute to the activation of cellulose biosynthesis through enhanced intracellular accumulation of the messenger molecule. We therefore developed a model for the molecular mechanism underlying the self-flocculation of *Z. mobilis* (Fig. 6). Cellulose is synthesized from UDP-glucose by the fusion protein encoded by ZMO1083/2, which is activated by the intracellular accumulation of c-di-GMP due to its impaired degradation caused by the mutation in ZMO1055; cellulose fibrils are subsequently synthesized and exported by the Bcs complex for the self-flocculation of the bacterial cells through their entanglement.
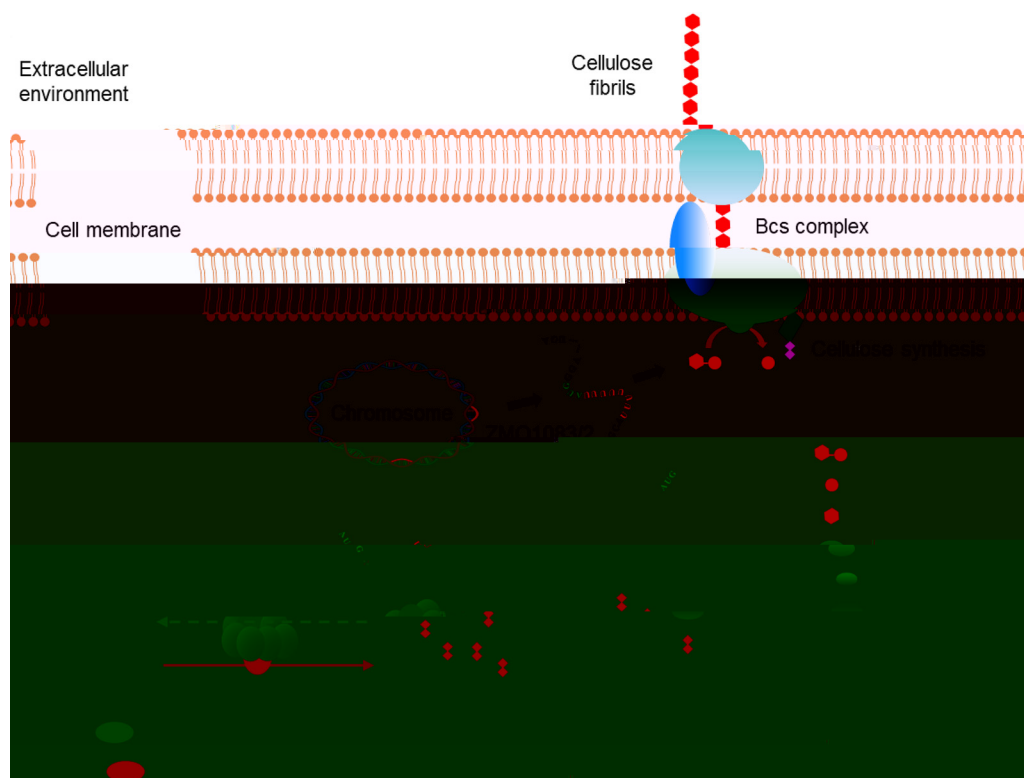
Cellulose production from glucose by

FIG 6 Mechanism underlying the self-flocculation of the bacterial cells (ZM401).

The structure of BcsA deciphered in *Rhodobacter sphaeroides* formed eight transmembrane helices for functions on catalyzing cellulose biosynthesis (43). BcsA in ZM4 contains 665 amino acids, and topological analysis of the protein structure indicates that six transmembrane helices could develop. With ZMO1082 fused with ZMO1083 in ZM401 due to the frameshift mutation, the gene ZMO1083/2 codes for a protein with 725 amino acids for two additional transmembrane helices at the N terminus, which besides altering binding of the product UDP, might interact with BcsB to promote cellulose biosynthesis or stabilize the protein, but whether binding of the substrate UDP-glucose is also altered needs to be studied. The two additional transmembrane helices thus increase the functionality of BcsA in ZM401 for cellulose biosynthesis and the development of cellulose fibrils as well for the self-flocculation of the bacterial cells. With 725 amino acids, BcsA in ZM401 resembles BcsA in *R. sphaeroides* with 788 amino acids (43). Both proteins can catalyze cellulose biosynthesis, and thus, deciphering the structure of BcsA in ZM401 to further explore the mechanism underlying catalysis for the biosynthesis of cellulose fibrils to flocculate the bacterial cells will unravel communal and distinct aspects of cellulose biosynthesis in *Z. mobilis*.

Another impact of the self-flocculation of ZM401 comes from ZMO1055. Although the SNP mutation with ZMO1055 in the genome of ZM401 did not substantially change the structure of this bifunctional GGDEF-EAL protein, the mutation might induce less stable binding of c-di-GMP with the phosphodiesterase EAL domain, and consequently compromise c-di-GMP degradation for intracellular accumulation of the messenger molecule to activate cellulose biosynthesis. The changes of intracellular c-di-GMP concentration in ZM401 engineered with the deletion of ZM1055 as well as the overexpression of ZMO1055 from ZM4 and its own ZMO1055 with the mutation experimentally supported such a speculation. Bioinformatic analysis and molecular docking showed that the mutation is not located in a consensus motif identified to be required for PDE activity, but compromises the binding of

c-di-GMP with the active site for degradation. Therefore, the EAL domain of the phosphodiesterase encoded by ZMO1055 with the SNP mutation regulates the cellulose biosynthesis by the Bcs complex in ZM401.

Although no significant difference in cell growth and ethanol fermentation was observed between ZM401 and ZM4 (Fig. S8), the self-flocculation of the bacterial cells should be controlled properly for industrial applications to avoid potential risk in mass transfer limitation on nutrient uptake and product diffusion caused by their strong self-flocculation to form large flocs. With the elucidation of the molecular mechanism underlying the self-flocculation of ZM401, strategies can be developed for engineering other unicellular strains of *Z. mobilis* with the self-flocculating phenotype, and in the meantime the self-flocculating process can be optimized at the molecular level. Moreover, theoretically and technically, other bacteria can also be engineered with the phenotype for robust production.

Conclusions. Due to the ED pathway for glycolysis with less biomass accumulation for high product yield, *Z. mobilis* has been acknowledged as a chassis to be engineered as microbial cell factories for biorefineries to produce bulk commodities such as biofuels and bio-based chemicals with major costs from feedstock consumption. The self-flocculation of *Z. mobilis* endows a unique morphology for the bacterial cells to develop multicellular flocs with physiological and metabolic merits to tolerate various stresses. With the elucidation of the molecular mechanism underlying such a phenotype, strategies can be developed for optimizing the self-flocculation of the bacterial cells through rational design at the molecular level for robust production.

MATERIALS AND METHODS

Strains and media. The strains used in this work are listed in Table S4. Rich media RMG2 and RMG10 containing 10 g/L yeast extract and 2 g/L KH_2PO_4 supplemented with 20 g/L and 100 g/L glucose, respectively, were used for seed culture and ethanol fermentation with *Z. mobilis*. Luria-Bertani (LB) medium was used for cultivating *E. coli* DH5α and *E. coli* JM110. All media were sterilized by autoclaving at 115°C for 20 min before inoculation.

Culture of *Z. mobilis* and ethanol fermentation. A single colony of ZM401 was inoculated into 5 mL RMG2, which was cultured at 30°C and 150 rpm for 24 h, and then used for DNA extraction and genome sequencing. A loopful of ZM401 or ZM4 was inoculated into an 150-mL Erlenmeyer flask containing 100 mL RMG2 for culture at 30°C and 150 rpm until their optical density at 600 nm (OD_{600}) increased to ~1.0 as seed for inoculation. The fermenter (2.5 L; KoBioTech KF, South Korea) containing 1.4 L RMG10 was inoculated with 100 mL seed culture for a total volume of 1.5 L, and ethanol fermentation was performed at 30°C and 150 rpm with pH controlled at 6.0 through automatic titration by 2 N KOH for a consistency of the fermentation process with the transcription analysis previously reported for ZM4 (44).

We observed that microaeration benefits the self-flocculation of ZM401 without significant impact on cell growth and ethanol production. Therefore, air filtered with a 0.2 μm-membrane was sparged into the fermenter at 100 mL/min to stimulate the self-flocculation of ZM401 for mining more information on gene expression associated with the self-flocculating phenotype of the bacterial cells.

Analysis of cell growth and ethanol fermentation. The cell suspension of 4 mL was sampled and deflocculated by cellulases at room temperature following the protocol developed previously (27), and was then vortexed vigorously for homogenous suspension, followed by OD_{600} measurement using a microplate reader (Multiskan GO; ThermoFisher, USA) to characterize the cell density.

Glucose, ethanol, and glycerol were analyzed by high-pressure liquid chromatography (HPLC) (Waters e2695, USA) with the refractive index (RI) detector (Waters 2414, USA) operated at 50°C. An organic acid column (Aminex HPX-87H, 300 × 7.8 mm; Bio-Rad, USA) was employed for the analysis, which was operated at 65°C with 10 mmol/L H_2SO_4 as the flow phase fed at a flow rate of 0.6 mL/min.

***In situ* characterization of the self-flocculation of *Z. mobilis*.** Bacterial flocs are deformable and not uniform in size, which cannot be characterized through microscopic observation, and thus, *in situ* measurement is needed. The one-dimensional chord length detected *in situ* by the FBRM system (ParticleTrack G400; Mettler Toledo, USA) was employed as previously reported for the measurement of the self-flocculating yeast (45). Bacterial flocs were collected and transferred into the container equipped with an impeller for the FBRM analysis, and the impeller was adjusted to a rotating rate of 150 rpm for homogenous suspension of the bacterial flocs. The beam from the FBRM sensor scanned the sample at a speed of 2 m/s with an interval of 10 s, and the process was run for at least 15 min with each sample to make the data set statistically reliable. Data were analyzed and processed using Macro v. 1.1.11. Square weight mean was used to assess the statistical data.

Genomic DNA extraction and genome sequencing. The genomic DNA was extracted from 5 mL culture inoculated with the single colony of ZM401 using the bacterial DNA kit (Omega, USA), following the protocol provided by the manufacturer.

The genome sequencing was performed using the third generation sequencing PacBio RS II system at Instrumental Analysis Center, Shanghai Jiao Tong University (Shanghai, China). The next generation sequencing (NGS) was performed using the Illumina X 10 platform at Genewiz, Inc. (Suzhou, China) with paired-end sequencing technology according to the standard protocols.

The SNVs of ZM401 were identified with the genome of ZM4 as the reference (GenBank accession numbers NZ_CP023715 for the chromosome and NZ_CP023716, NZ_CP023717, NZ_CP023718, and NZ_CP023719 for the plasmids). The structure variations (SVs) were analyzed with the long-read PacBio data based on previously reported methods (46).

Transcriptomic analysis. The transcriptomic analysis workflow was developed based on previous work (44). Briefly, cultures were sampled from the fermenter at 6 h when the bacterial cells grew to their exponential stage, followed by total RNA extraction using the TRIzol reagent (Invitrogen, USA). rRNA was digested using the Ribo-off rRNA depletion kit (Vazyme, China). The sequencing library was then constructed after connecting the sequencing adapters with the cDNA fragments, of which the first-strand cDNA was synthesized using random hexamer-primers, and the second-strand cDNA was synthesized using dATPs, dGTPs, dCTPs, dUTPs, RNase H, and DNA polymerase I after removing deoxynucleoside triphosphates (dNTPs).

RNA-seq was performed at Genewiz, Inc. (Suzhou, China) using paired-end sequencing technology following the manufacturer's protocol. Data quality was assessed with the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and data passing the quality control process were imported into the CLC Genomics Workbench v.11.0 (Qiagen, USA) for analysis to obtain the RPKM (reads mapping to the genome per kilobase of transcript per million reads sequenced) values using the genome of ZM4 as the reference. Gene expression normalization, analysis of variance, and hierarchical clustering development were conducted using JMP Genomics v. 9.0 (SAS, Inc., USA) to identify differentially expressed genes in ZM401 with the selection thresholds $-1 \geq \log_2$ (fold change) $\geq +1$ and the statistical significance of $P \leq 0.05$. Duplicate samples were used for each analysis.

Protein-ligand docking. With crystal structures of BcsA from *R. sphaeroides* (PDB ID: 4p02.1.A) and PA0861 from *P. aeruginosa* (PDB ID: 5xge.1) as templates for simulating BcsA and ZMO1055 in ZM4 and ZM401, respectively (38, 47), protein structures were constructed using the protein homology modeling software SWISS-MODEL (48). Global model quality estimation (GMQE) and qualitative model energy analysis (QMEAN) were applied to assess the modeling results, and a criterion with scores between 0 and 1 was recommended for such an assessment (48). For BcsA and ZMO1055 in ZM4 and ZM401, their GMQE and QMEAN scores were calculated to be 0.64 and 0.62 as well as 0.62 and 0.63, respectively, which are statistically significant for simulations using the model. Then, substrates and activators were docked with enzymes using the software BIOVIA Discovery Studio 2016 to simulate their interactions, and binding energy was further calculated.

Genetic manipulations. The plasmids and primers used in this study are listed in Tables S5 and S6, respectively. Plasmid construction and genetic manipulations were performed as previously reported (26). For gene deletion, two homologous regions with the target gene flanked were obtained by PCR, and then infused with the plasmid pEX18Tc. For gene overexpression, target genes and promoters were amplified by PCR, and then infused with the plasmid pHW20a or pEZ15Asp. Recombinant plasmids were transformed into *E. coli* DH5α for propagation, which were confirmed by Sanger sequencing before being transformed into *E. coli* JM110 for demethylation. Demethylated recombined plasmids were then transformed into ZM401 or ZM4 through electroporation using Gene Pulser (Bio-Rad, USA) with 0.1-cm gap cuvettes operated at 18 kV/cm, 200 Ω, and 25 μF. Selection of positive colonies was carried out on RMG2 agar plates containing 20 mg/L tetracycline or 100 mg/L spectinomycin for mutants engineered with the vector pHW20a or vector pEZ15Asp. While engineered with the vector pEX18Tc, colonies were screened on RMG2 agar plates supplemented with 20 mg/L tetracycline followed by another selection on RM agar plates composed of 10 g/L yeast extract and 2 g/L KH$_2$PO$_4$ and 50 g/L sucrose but without glucose. PCR verification was performed (Fig. S9).

Extraction and measurement of intracellular c-di-GMP. The bacterial cells were grown for 12 h to their exponential phase with an OD$_{600}$ ~3.0, and then 15 mL culture was centrifuged at 5,000 × *g* and 4°C for 3 min to collect the pellet. The precooled solvent of 1 mL (2:2:1 for methanol:acetonitrile:double-distilled water [ddH$_2$O]) was used to disrupt the bacterial cells and extract soluble components at −20°C for 30 min, and the supernatant was collected by centrifugation at 12,000 × *g* and 4°C for 5 min. The pellet was dispersed again with 1 mL precooled extraction solvent for another round of extraction and centrifugation with supernatant collected.

Chloroform (2 mL) was added into the supernatant (2 mL), which was mixed by shaking to extract the hydrophobic components selectively. After resting 5 min for phase separation, the upper phase with c-di-GMP was collected for drying by a freeze dryer. The extraction solvent of 80 μL was used to dissolve the paste, which was vortexed violently for 2 min, and then 50 μL solution was sampled for c-di-GMP analysis with an Acquity ultra performance liquid chromatography (UPLC) I-class/Vion IMS QToF instrument.

Briefly, an I-class Acquity UPLC (Waters, UK) instrument with a SeQuant ZIC-HILIC column (100 × 2.1 mm) packed with 3.5 μm polyetheretherketone was used for separating c-di-GMP at 45°C. The flow phase was developed by phase A (50 mM ammonium formate) and phase B (acetonitrile). Separation of c-di-GMP was conducted at three stages within 15 min at a flow rate of 0.4 mL/min as follows: stage I from 0 to 10 min with phase B decreased from 90% to 50% and phase A increased from 10% to 50%, stage II from 10 min to 12 min with phase B increased from 50% to 90% and phase A decreased from 50% to 10%, and stage III from 12 min to 15 min with 90% phase B plus 10% phase A. c-di-GMP was further analyzed with a Vion IMS QToF mass spectrometer (Waters, UK) with an electrospray ionization (ESI) interface operated at negative ion mode with the parameters desolvation and ion source temperatures of 500 and 120°C, respectively; collision energy, 40 eV; desolvation gas flow rate, 1,000 L/h; cone gas flow rate, 50 L/h; capillary voltage, 2,000 V; cone voltage, 20 V. Data were acquired from *m/z* 50 to 1,000 at a scan speed of 0.3 s with the scan mode mRm. UNIFI v. 1.8.1 was used for data processing, and c-di-GMP with a purity of 98% (Biolog, Germany) was used as the standard to calibrate the analysis.

Data availability. All data are available in the main text and supplemental material. The raw data for omics analysis were deposited at the National Center for Biotechnology Information (NCBI) with the

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, MP4 file, 0.7 MB.
SUPPLEMENTAL FILE 2, MP4 file, 0.8 MB.
SUPPLEMENTAL FILE 3, PDF file, 1.4 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S. 2016. Biofilms: an emergent form of bacterial life. Nat Rev Microbiol 14: 563–575. https://doi.org/10.1038/nrmicro.2016.94.
2. Yan J, Bassler BL. 2019. Surviving as a community: antibiotic tolerance and persistence in bacterial biofilms. Cell Host Microbe 26:15–21. https://doi.org/10.1016/j.chom.2019.06.002.
3. Mukherjee S, Bassler BL. 2019. Bacterial quorum sensing in complex and dynamically changing environments. Nat Rev Microbiol 17:371–382. https://doi.org/10.1038/s41579-019-0186-5.
4. Karygianni L, Ren Z, Koo H, Thurnheer T. 2020. Biofilm matrixome: extracellular components in structured microbial communities. Trends Microbiol 28:668–681. https://doi.org/10.1016/j.tim.2020.03.016.
5. Rumbaugh KP, Sauer K. 2020. Biofilm dispersion. Nat Rev Microbiol 18: 571–586. https://doi.org/10.1038/s41579-020-0385-0.
6. Wilén BM, Liébana R, Persson F, Modin O, Hermansson M. 2018. The mechanisms of granulation of activated sludge in wastewater treatment, its optimization, and impact on effluent quality. Appl Microbiol Biotechnol 102:5005–5020. https://doi.org/10.1007/s00253-018-8990-9.
7. Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. 2016. The activated sludge ecosystem contains a core community of abundant organisms. ISME J 10:11–20. https://doi.org/10.1038/ismej.2015.117.
8. Saha S, Basak B, Hwang JH, Salama ES, Chatterjee PK, Jeon BH. 2020. Microbial symbiosis: a network towards biomethanation. Trends Microbiol 28:968–984. https://doi.org/10.1016/j.tim.2020.03.012.
9. van Loosdrecht MC, Brdjanovic D. 2014. Water treatment: anticipating the next century of wastewater treatment. Science 344:1452–1453. https://doi.org/10.1126/science.1255183.
10. Zhao XQ, Bai FW. 2009. Yeast flocculation: new story in fuel ethanol production. Biotechnol Adv 27:849–856. https://doi.org/10.1016/j.biotechadv.2009.06.006.
11. Mukhopadhyay A. 2015. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. Trends Microbiol 23: 498–508. https://doi.org/10.1016/j.tim.2015.04.008.
12. Sprenger GA. 1996. Carbohydrate metabolism in *Zymomonas mobilis*: a catabolic highway with some scenic routes. FEMS Microbiol Lett 145: 301–307. https://doi.org/10.1111/j.1574-6968.1996.tb08593.x.
13. Rutkis R, Strazdina I, Balodite E, Lasa Z, Galinina N, Kalnenieks U. 2016. The low energy-coupling respiration in *Zymomonas mobilis* accelerates flux in the Entner-Doudoroff pathway. PLoS One 11:e0153866. https://doi.org/10.1371/journal.pone.0153866.
14. Wang X, He Q, Yang Y, Wang J, Haning K, Hu Y, Wu B, He M, Zhang Y, Bao J, Contreras LM, Yang S. 2018. Advances and prospects in metabolic engineering of *Zymomonas mobilis*. Metab Eng 50:57–73. https://doi.org/10.1016/j.ymben.2018.04.001.
15. Liu Y, Ghosh IN, Martien J, Zhang Y, Amador-Noguez D, Landick R. 2020. Regulated redirection of central carbon flux enhances anaerobic production of bioproducts in *Zymomonas mobilis*. Metab Eng 61:261–274. https://doi.org/10.1016/j.ymben.2020.06.005.
16. Xia J, Yang YF, Liu CG, Yang SH, Bai FW. 2019. Engineering *Zymomonas mobilis* for robust cellulosic ethanol production. Trends Biotechnol 37: 960–972. https://doi.org/10.1016/j.tibtech.2019.02.002.
17. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. Curr Opin Biotechnol 45:202–211. https://doi.org/10.1016/j.copbio.2017.03.008––

NF, Brown SD. 2009. Improved genome annotation for *Zymomonas mobilis*. Nat Biotechnol 27:893–894. https://doi.org/10.1038/nbt1009-893.

30. Yang S, Vera JM, Grass J, Savvakis G, Moskvin OV, Yang Y, McIlwain SJ, Lyu Y, Zinonos I, Hebert AS, Coon JJ, Bates DM, Sato TK, Brown SD, Himmel ME, Zhang M, Landick R, Pappas KM, Zhang Y. 2018. Complete genome sequence and the expression pattern of plasmids of the model ethanologen *Zymomonas mobilis* ZM4 and its xylose-utilizing derivatives 8b and 2032. Biotechnol Biofuels 11:125. https://doi.org/10.1186/s13068-018-1116-x.

31. Mao F, Dam P, Chou J, Olman V, Xu Y. 2009. DOOR: a database for prokaryotic operons. Nucleic Acids Res 37:D459–63. https://doi.org/10.1093/nar/gkn757.

32. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613. https://doi.org/10.1093/nar/gky1131.

33. Omadjela O, Narahari A, Strumillo J, Mélida H, Mazur O, Bulone V, Zimmer J. 2013. BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. Proc Natl Acad Sci U S A 110:17856–17861. https://doi.org/10.1073/pnas.1314063110.

34. Salgado L, Blank S, Esfahani RAM, Strap JL, Bonetta D. 2019. Missense mutations in a transmembrane domain of the *Komagataeibacter xylinus* BcsA lead to changes in cellulose synthesis. BMC Microbiol 19:216. https://doi.org/10.1186/s12866-019-1577-5.

35. Jenal U, Reinders A, Lori C. 2017. Cyclic di-GMP: second messenger extraordinaire. Nat Rev Microbiol 15:271–284. https://doi.org/10.1038/nrmicro.2016.190.

36. Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, Benziman M. 1987. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. Nature 325:279–281. https://doi.org/10.1038/325279a0.

37. Nesbitt NM, Arora DP, Johnson RA, Boon EM. 2015. Modification of a bifunctional diguanylate cyclase-phosphodiesterase to efficiently produce cyclic diguanylate monophosphate. Biotechnol Rep (Amst) 7:30–37. https://doi.org/10.1016/j.btre.2015.04.008.

38. Morgan JL, McNamara JT, Zimmer J. 2014. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. Nat Struct Mol Biol 21:489–496. https://doi.org/10.1038/nsmb.2803.

39. Jeon YJ. 2019. Recent progress in strain development of *Zymomonas mobilis* for lignocellulosic ethanol production. J Life Sci 29:135–145. https://doi.org/10.5352/JLS.2019.29.1.135.

40. Yang S, Mohagheghi A, Franden MA, Chou YC, Chen X, Dowe N, Himmel ME, Zhang M. 2016. Metabolic engineering of *Zymomonas mobilis* for 2,3-butanediol production from lignocellulosic biomass sugars. Biotechnol Biofuels 9:189. https://doi.org/10.1186/s13068-016-0606-y.

41. Römling U, Galperin MY. 2015. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. Trends Microbiol 23:545–557. https://doi.org/10.1016/j.tim.2015.05.005.

42. Jacek P, Dourado F, Gama M, Bielecki S. 2019. Molecular aspects of bacterial nanocellulose biosynthesis. Microb Biotechnol 12:633–649. https://doi.org/10.1111/1751-7915.13386.

43. Morgan JL, Strumillo J, Zimmer J. 2013. Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature 493:181–186. https://doi.org/10.1038/nature11744.

44. Yang S, Tschaplinski TJ, Engle NL, Carroll SL, Martin SL, Davison BH, Palumbo AV, Rodriguez M, Jr, Brown SD. 2009. Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. BMC Genomics 10:34. https://doi.org/10.1186/1471-2164-10-34.

45. Ge XM, Zhao XQ, Bai FW. 2005. Online monitoring and characterization of flocculating yeast cell flocs during continuous ethanol fermentation. Biotechnol Bioeng 90:523–531. https://doi.org/10.1002/bit.20391.

46. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 15:461–468. https://doi.org/10.1038/s41592-018-0001-7.

47. Liu C, Liew CW, Wong YH, Tan ST, Poh WH, Manimekalai MSS, Rajan S, Xin L, Liang ZX, Grüber G, Rice SA, Lescar J. 2018. Insights into biofilm dispersal regulation from the crystal structure of the PAS-GGDEF-EAL region of RbdA from *Pseudomonas aeruginosa*. J Bacteriol 200:e00515-17. https://doi.org/10.1128/JB.00515-17.

48. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303. https://doi.org/10.1093/nar/gky427.