Supplementary materials for

Construction and optimization of a microbial platform for sustainable biosynthesis of poly-*N*-acetyllactosamine glycoprotein in the cytoplasm for detecting tumor biomarker galectin-3

Jian Yuan[a,b], Xiaolong Jiang[a,b], Huijing Zhang[a,b], Jing Bai[a,b], Jing Zhang[c,d], Yuhui Wang[a,b], Gege Lu[a,b], Huimin Xu[c,d], Huanhuan Liu[c,d,*], Bin Liu[a,b,*], Di Huang[a,b,*]

[a] Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, Nankai University, Tianjin, China

[b] TEDA Institute of Biological Sciences and Biotechnology, Tianjin Key Laboratory of Microbial Functional Genomics, Nankai University, Tianjin, China

[c] State Key Laboratory of Food Nutrition and Safety, Tianjin University of Science & Technology, Tianjin, China.

[d] Key Laboratory of Food Nutrition and Safety, Ministry of Education, Tianjin University of Science & Technology, Tianjin, China.

Correspondence author:

[*]Address for Correspondence: TEDA Institute of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, Tianjin Economic-Technological Development Area (TEDA), Tianjin, 300457, P R China. Tel: 86-22-6622 9592. Fax: 86-22-6622 9596. E-mail: huangdi@nankai.edu.cn or liubin1981@nankai.edu.cn or lh_tust@tust.edu.cn

Jian Yuan and Xiaolong Jiang contributed equally to this work.

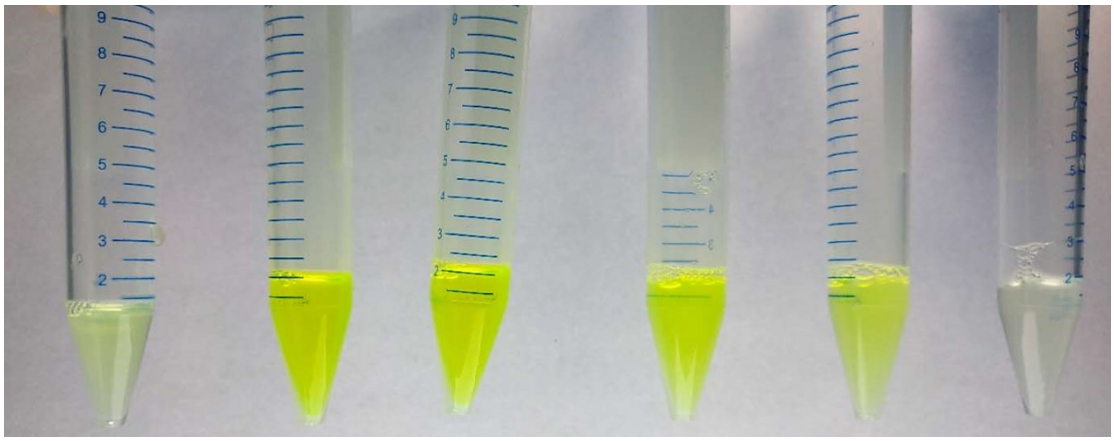
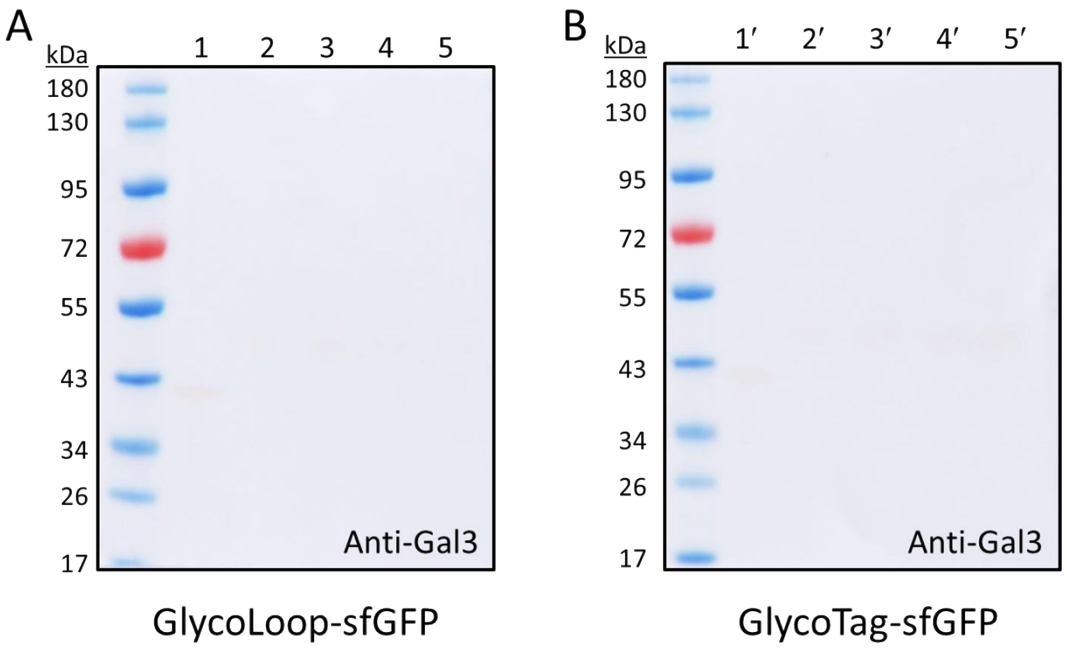


Fig. S1 Purified glycosylated sfGFP.



1: JM811(JM109/GlycoLoop-sfGFP)
2: JM812(JM109/GlycoLoop-sfGFP-GtfAB)
3: JM813(JM109/GlycoLoop-sfGFP-GtfAB-LgtB)
4: JM814(JM109/GlycoLoop-sfGFP-GtfAB-LgtB-LgtA)
5: JM815(JM109/GlycoLoop-sfGFP-GtfAB-LgtB-LgtA/galUE-pgm)

1′: JM821(JM109/GlycoTag-sfGFP)
2′: JM822(JM109/GlycoTag-sfGFP-GtfAB)
3′: JM823(JM109/GlycoTag-sfGFP-GtfAB-LgtB)
4′: JM824(JM109/GlycoTag-sfGFP-GtfAB-LgtB-LgtA)
5′: JM825(JM109/GlycoTag-sfGFP-GtfAB-LgtB-LgtA/galUE-pgm)

Fig. S2 Analysis of sfGFP glycosylation with poly-LacNAc in the cytoplasm of *E. coli* JM109. Immunoblot analysis using anti-Gal3 antibody was performed to detect GlycoLoop-sfGFP glycoproteins (A) and GlycoTag-sfGFP glycoproteins (B) purified from the cytoplasm fraction of *E. coli* JM109.

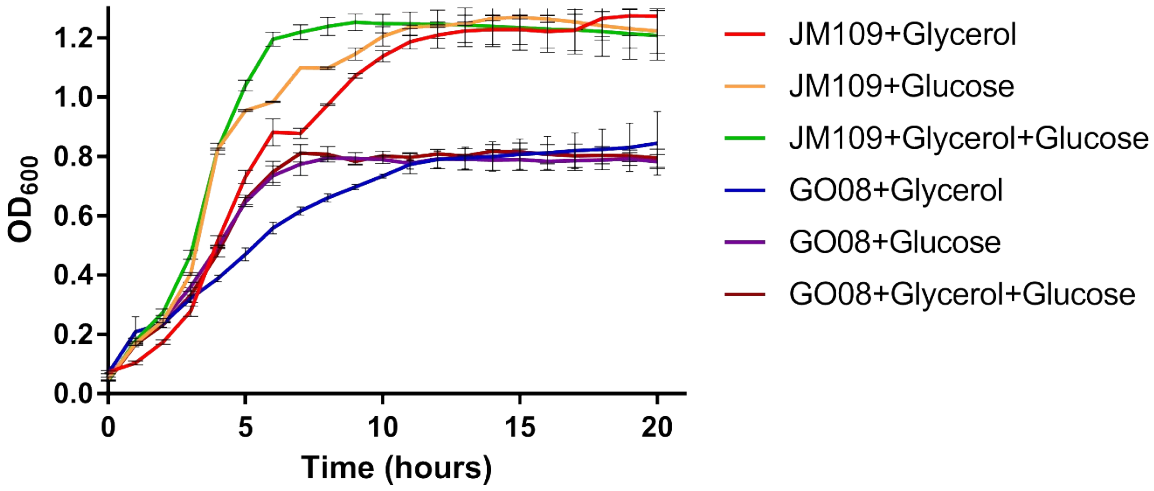
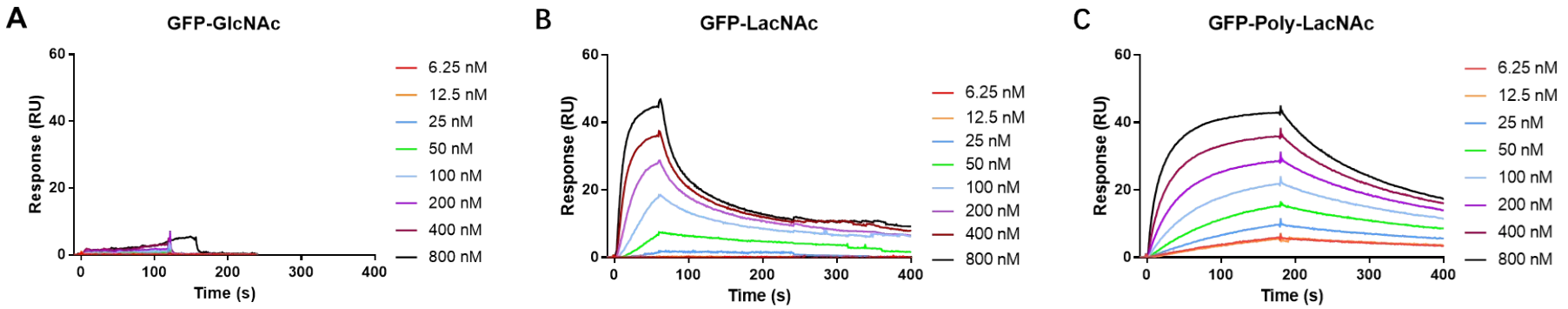


Fig. S3 Growth profiles of *E. coli* K-12 JM109 and chassis strain GO08 cultivated in glucose or (and) glycerol M9Y medium.



D

Analytes	Association constant (k_a, $M^{-1} \cdot s^{-1}$)	Dissociation constant (k_d, s^{-1})	Binding constant (K_D, M)
GFP	-	-	-
GFP-GlcNAc	7.727×10^5	8.173×10^{-1}	1.058×10^{-6}
GFP-LacNAc	1.011×10^7	4.462×10^{-2}	4.414×10^{-9}
GFP-Poly-LacNAc	1.018×10^6	6.634×10^{-2}	6.518×10^{-8}

Fig. S4 Analysis by SPR spectroscopy of the binding of glycoprotein to the galectin CGL2 immobilized on a BIAcore CM5 chip. Different concentrations (6.25, 12.5, 25, 50, 100, 200, 400 and 800 nM) of GFP-GlcNAc (A), GFP-LacNAc (B), GFP-poly-LacNAc (C) in HEPES buffered saline were injected at a flow rate of 50 µL/min. (D) Kinetic rate constants (k_a and k_d) were fitted using the Bivalent-Analyte binding model. The binding constant K_D was calculated by dividing k_d by k_a.

Table S1 Strains and plasmids used in this study.

Strains/Plasmids	Relevant characteristics	Reference or source
Strains		
DH5α	F$^-$ *glnV44 deoR nupG purB20* Φ80*dlacZ*ΔM15Δ(*lacZYA-argF*)U169 *recA1 endA1 hsd*R17(*rk*$^-$, *mk*$^+$) *thi*-1 *gyr*A96 *relA1* λ−	Lab stock
JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB*$^+$ Δ(*lac-proAB*) e14- [F' *traD36 proAB*$^+$ *lacIq lacZ*ΔM15] *hsdR17*($r_K^-m_K^+$)	Lab stock
BL21(DE3)	F$^-$*ompT hsdS$_B$* ($r_B^-m_B^-$) *gal dcm rne131* λ(DE3)	Lab stock
JM811	JM109/pGlycoLoop-sfGFP	This study
JM812	JM109/pGlycoLoop-sfGFP-GtfAB	This study
JM813	JM109/pGlycoLoop-sfGFP-GtfAB-LgtB	This study
JM814	JM109/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA	This study
JM815	JM109/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA/pCDFm-GalU-GalE-Pgm	This study
JM821	JM109/pGlycoTag-sfGFP	This study
JM822	JM109/pGlycoTag-sfGFP-GtfAB	This study
JM823	JM109/pGlycoTag-sfGFP-GtfAB-LgtB	This study
JM824	JM109/pGlycoTag-sfGFP-GtfAB-LgtB-LgtA	This study
JM825	JM109/pGlycoTag-sfGFP-GtfAB-LgtB-LgtA/pCDFm-GalU-GalE-Pgm	This study
GO01	JM109 Δ*lacIZYA*	This study
GO02	JM109 Δ*lacIZYA* Δ*zwf*	This study
GO03	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB*	This study
GO04	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB* Δ*pfkA*	This study

GO05	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB* Δ*pfkA* Δ*pykF*	This study
GO06	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB* Δ*pfkA* Δ*pykF* Δ*pykA*	This study
GO07	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB* Δ*pfkA* Δ*pykF* Δ*pykA* Δ*gldA*	This study
GO08	JM109 Δ*lacIZYA* Δ*zwf* Δ*pfkB* Δ*pfkA* Δ*pykF* Δ*pykA* Δ*gldA* Δ*nagB*	This study
GO811	GO08/pGlycoLoop-sfGFP	This study
GO812	GO08/pGlycoLoop-sfGFP-GtfAB	This study
GO813	GO08/pGlycoLoop-sfGFP-GtfAB-LgtB	This study
GO814	GO08/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA	This study
GO815	GO08/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA/pCDFm-GalU-GalE-Pgm	This study
GO816	GO08/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA/pET28m-GlmS54-GlmM-GlmU	This study
GO817	GO08/pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA/pCDFm-GalU-GalE-Pgm/pET28m-GlmS54-GlmM-GlmU	This study
GO821	GO08/pGlycoTag-sfGFP	This study
GO822	GO08/pGlycoTag-sfGFP-GtfAB	This study
GO823	GO08/pGlycoTag-sfGFP-GtfAB-LgtB	This study
GO824	GO08/pGlycoTag-sfGFP-GtfAB-LgtB-LgtA	This study
Plasmids		
pACYCDuet-1	Double T7 promoters; P15A ori Cm^R	Novagen
pCDFDuet-1	Double T7 promoters; CloDF13 ori $Spec^R$	Novagen
pET28a (+)	Single T7 promoter; ColE ori Kan^R	Novagen
pACYCm	pACYCDuet-1 derived plasmid, the original double T_7 promoters were replaced by double Ptac promoters; P15A ori Cm^R	This study
pGlycoTag-sfGFP	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and a C terminally fused glycosylation sequence	This study
pGlycoLoop-sfGFP	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and an internal loop	This study

	fused glycosylation sequence	
pGlycoTag-sfGFP-GtfAB	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and a C terminally fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213	This study
pGlycoLoop-sfGFP-GtfAB	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and an internal loop fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213	This study
pGlycoTag-sfGFP-GtfAB-LgtB	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and a C terminally fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213, and codon-optimized *lgtB* from *N. meningitidis* MC58	This study
pGlycoLoop-sfGFP-GtfAB-LgtB	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and an internal loop fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213, and codon-optimized *lgtB* from *N. meningitidis* MC58	This study
pGlycoTag-sfGFP-GtfAB-LgtB-LgtA	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and a C terminally fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213, and codon-optimized *lgtB* from *N. meningitidis* MC58, and codon-optimized *lgtA* from *H. pylori* J99	This study
pGlycoLoop-sfGFP-GtfAB-LgtB-LgtA	pACYCm carrying codon-optimized *sfGFP* with a N terminally fused hexahistidine tag and an internal loop fused glycosylation sequence, and codon-optimized *gtfA-gtfB* from *S. parasanguinis* FW213, and codon-optimized *lgtB* from *N. meningitidis* MC58, and codon-optimized *lgtA* from *H. pylori* J99	This study
pCDFm	pCDFDuet-1 derived plasmid, the original double T_7 region between EcoNI and XhoI site was replaced by a 1333-nucleotide fragment containing double Ptac promoters, rrnB_T1/T2 terminators, followed by the insertion of double multiple cloning sites between the promoter and terminator; CloDF13 ori $Spec^R$	This study
pCDFm-GalU-GalE	pCDFm carrying *galU* and *galE* from *E. coli*	This study
pCDFm-GalU-GalE-Pgm	pCDFm carrying *galU*, *galE* and *pgm* from *E. coli*	This study
pET28m	pET28a (+) derived plasmid, the original T_7 region between BglII and XhoI site was replaced by a 1236-nucleotide fragment containing double Ptac promoter, rrnB_T1/T2 terminator; ColE ori Kan^R	This study
pET28m-GlmS54	pET28m carrying codon-optimized *glmS*54* from *E. coli*	This study

pET28m-GlmS54-GlmM-GlmU	pET28m carrying codon-optimized *glmS*54*, *glmM* and *glmU* from *E. coli*	This study
pKD4	*FRT-kan-FRT bla* R6K γ	[1]
pKD46	*rep*A101(Ts) *pSC101 bla ara*C P_{araBAD}-*exo*, *beta* and *gama*	[1]
pCas	*rep*A101(Ts) *pSC101 kan Pcas-cas*9 P_{araBAD}- *exo*, *beta* and *gama lac*I^{q} P_{trc}-sgRNA-*pMB1*	[2]
pTarget	*pMB1 aadA*	[2]
pCP20	*rep*A101(Ts) *pSC101 bla cat* λ repressor (ts) FLP	[3]

Table S2 N20 sequences of gRNA used for CRISPR-Cas9 based genome editing used in this study.

Gene	N20 Sequence at the 5' end of protospacer-adjacent motif (PAM)
lacIZYA	CGATGGTTTGCCCGGATAAA
zwf	GATCATGTCGCGCATCTGAC
pfkA	TGAAACCCGCGCAACTGTGC
pfkB	AGTGCAGCCACACTCAATCA
nagB	CGAAATGCATAAAGCAGGCC
pykF	CACCGGCTTCTGCGCGAGTC
pykA	TCCATGACTTCTGCACGCGT
gldA	TATTTGGCAACAGCAGATAG

Materials and Methods

Bacterial strains and growth conditions

The strains used in this study are listed in Supplementary material Table S1. *E. coli* DH5α was used for the maintenance and propagation of plasmids, while *E. coli* BL21(DE3) was applied to express CGL2 protein. Both DH5α and BL21(DE3) were cultured at 37 °C in Luria-Bertani (LB) medium containing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl in shake flasks, or on LB plates. *E. coli* K-12 JM109 and established glycoengineered strain GO08 were used as hosts for the production of glycoprotein. Cultures were grown overnight in LB medium (20 mL) containing appropriate antibiotics to prepare a seed culture. Seed culture (1 mL) was re-inoculated into the production medium (100 mL) in a shake flask. Production was carried out with the modified M9 (M9Y) medium as described previously,[4] containing 20 g/L glycerol, 20 g/L glucose, 6 g/L Na_2HPO_4, 0.5 g/L NaCl, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 2 g/L yeast extract. When needed, 50 mg/L of kanamycin, 25 mg/L of chloramphenicol, and 50 mg/L of spectinomycin were supplemented. Cultures were grown at 37 °C to an absorbance at 600 nm (optical density OD_{600}) of 0.6–0.8. Expression was induced by adding various concentrations (0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM, and 1 mM) of inducer isopropyl β-D-1-thiogalactopyranoside

(IPTG). Cultures were grown for a further 12–72 h at various induction temperatures (16 °C, 20 °C, 28 °C, 34 °C, and 37 °C) in the presence of inducing agents.

Plasmid construction

The plasmids and primers used in this study are listed in Supplementary material Table S1. The vectors pACYCDuet-1, pCDFDuet-1, and pET28a(+) (Novagen) were used for expression of glycosylated superfolder green fluorescent protein (sfGFP), the precursor pathway, and CGL2 galectin, respectively. Molecular biology procedures followed standard protocols as described previously.[5]

The plasmid pACYCm was modified from pACYCDuet-1 by replacing the original T7 region between *Eco*NI and *Xho*I by an approximately 1000-nucleotide fragment containing double tac promoters, double rrnB terminators, and double multiple cloning sites. The sfGFP with an N-terminally fused hexahistidine tag, which incorporated the glycosylation sequence in a C-terminal tag or an internal loop (between residues 194 and 195), were codon-optimized and synthesized by General Biosystems (Anhui, China) and cloned into *Eco*NI and *Bam*HI of pACYCDuet-1 (Novagen) to remove native T7 promoters, resulting in pGlycoTag-sfGFP or pGlycoLoop-sfGFP. Subsequently, the *gtfA-gtfB* that initiates the first step of the protein glycosylation by transferring GlcNAc to the Ser residues of the protein backbone was derived from the genomic DNA of *S. parasanguinis* FW213 and synthesized after codon optimization. The gene was digested with *Bam*HI and *Eco*RI and ligated into the corresponding sites of pGlycoTag-sfGFP or pGlycoLoop-sfGFP. Then, the codon-optimized *lgtB* gene encoding β-1,4-galactosyltransferase from *N. meningitidis* MC58 was cut with *Sal*I and *Not*I and inserted into pGlycoTag-sfGFP-GtfAB or pGlycoLoop-sfGFP-GtfAB to obtain pGlycoTag-sfGFP-GtfAB-LgtB or pGlycoLoop-sfGFP-GtfAB-LgtB. Moreover, the codon-optimized *lgtA* gene encoding β-1,3-*N*-acetylglucosaminyltransferase from *H. pylori* J99 was digested with *Eco*RI and *Sal*I and inserted into pGlycoTag-sfGFP-GtfAB-LgtB or pGlycoLoop-sfGFP-GtfAB-LgtB, generating pGlycoTag-sfGFP-GtfAB-LgtB-LgtA or pGlycoLoop-sfGFP-

GtfAB-LgtB-LgtA, respectively. Each of the above-constructed open-reading frames was fused with a tac promoter, a ribosome-binding site, and an rrnB terminator operon.

For the expression of fungal galectin CGL2, the gene encoding the fungal galectin CGL2 from *Coprinus cinereus* was codon-optimized. The PCR product, incorporating an N-terminally fused hexahistidine tag was digested with *Nde*I/*Not*I and ligated to pET28a(+), resulting in the expression plasmid pET28a-CcCGL2.

The plasmid pCDFm was modified from pCDFDuet-1 by replacing double T7 promoters and one T7 terminator with double tac promoters and double rrnB terminators, followed by the insertion of double multiple cloning sites between the promoter and terminator. Genes *pgm*, *galU*, and *galE* were amplified from *E. coli*. First, *galU* and *galE* were fused by overlap-extension PCR, digested with *Nde*I/*Not*I and ligated to the second multiple cloning site, yielding pCDFm-GalU-GalE. Then, *pgm* was inserted between *Eco*RI and *Hin*dIII of the first multiple cloning site, generating pCDFm-GalU-GalE-Pgm.

The plasmid pET28m was modified from pET28a (+) by replacing the original T7 region between *Bgl*II and *Xho*I by an approximately 1000-nucleotide fragment containing double tac promoters, double rrnB terminators, and double multiple cloning sites between the promoter and terminator. Genes *glmM* and *glmU* were amplified from *E. coli*, and *glmS*54* was codon-optimized and synthesized. First, *glmS*54* was inserted into the first multiple cloning site between *Eco*RI and *Hin*dIII, generating pET28m-GlmS54. Then, polycistron *glmM* and *glmU*, with their intergenic region, were ligated to the second multiple cloning site *Nde*I/*Not*I, yielding pET28m-GlmS54-GlmM-GlmU. The correct insertion of all genes was verified by sequencing.

Construction of *E. coli* mutant strains by gene deletion

Single gene deletion from the *E. coli* K-12 JM109 chromosome was performed using the CRISPR-Cas9 method, as described previously.[6] The pTarget series was constructed by assembling the pTarget backbone with a targeting N20 sequence of gene loci of interest and two homology arms. The pTarget backbone was amplified by

inverse PCR with the modified N20 sequence overhanging at the 5′ ends of primers and 15–20 bp overlapping ends. The donor DNA with 500-bp homologous arms on each side was designed according to the sequence of the target gene and the location of the sgRNA. Two homology arms were separately amplified, fused together by overlap-extension PCR, and the resulting fragment also had overhangs with 15–20 bp overlapping ends. The pTarget backbone with a targeting N20 sequence and donor DNA was assembled using the Gibson assembly method. The designed sgRNAs targeting *zwf*, *pfkAB*, *pykFA*, *gldA*, *nagB*, and *lacIZYA* genes are shown in Supplementary material Table S2.

JM109 was transformed by pCas, and 10 mM arabinose was added to induce the expression of λ-Red. Subsequently, electrocompetent cells were mixed with pTarget series DNA, suspended immediately in ice-cold LB medium (1 mL), recovered for 3 h at 30 °C before spreading onto LB agar plate containing kanamycin and spectinomycin, and incubated overnight at 30 °C. Transformants were verified by colony PCR and DNA sequencing. The pTarget plasmid was cured by transferring positive colonies into LB medium containing 0.5 mM IPTG to induce sgRNA expression of pCas and cultivating for 8–10 h at 30 °C. The pCas plasmid was eliminated by cultivating overnight at 42 °C. The cultures after plasmid curing were streaked, and the colonies were tested for kanamycin and spectinomycin sensitivity, and were verified using PCR and sequencing. The method was repeated to delete all candidate genes.

Analytical methods

Cell growth was monitored by measuring OD_{600}, which was then converted to dry cell weight (DCW) using a calibration curve. The concentrations of glycerol and glucose in the batch fermentations were quantified using high-performance liquid chromatography (HPLC) (Agilent 1200 Series) with a refractive index (RI) detector and an organic acid column (Phenomenex, Torrance, CA, USA). The column was eluted with 5 mM H_2SO_4 at a constant flow rate of 0.6 mL/min at 50 °C. The concentration of glycoprotein was determined using a Pierce™ BCA Protein Assay Kit

(Thermo Scientific™, San Jose, CA, USA), according to the manufacturer's instructions.

References

1 K. A. Datsenko and B. L. Wanner, *Proceedings of the National Academy of Sciences*, 2000, **97**, 6640-6645.

2 Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang and S. Yang, *Appl. Environ. Microbiol.*, 2015, **81**, 2506-2514.

3 P. P. Cherepanov and W. Wackernagel, *Gene*, 1995, **158**, 9-14.

4 Y. Wu, X. Sun, Y. Lin, X. Shen, Y. Yang, R. Jain, Q. Yuan and Y. Yan, *Metab. Eng.*, 2017, **39**, 1-8.

5 J. Sambrook and D. W. Russell, *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York, 2001.

6 Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang and S. Yang, *Appl. Environ. Microbiol.*, 2015, **81**, 2506-2514.