Supplementary materials for

Sustainable production of polysaccharide-based glycoprotein by simultaneous conversion of glucose and glycerol in engineered *Escherichia coli*

Yuhui Wang[a,b], Xiaohan Wang[a,b], Guozhen Ma[a,b], Lijie Xie[a,b], Dan Liu[a,b], Yanling Wang[a,b], Xinyu Zhao[a,b], Yingying Su[a,b],Andrei V. Perepelov[c], Peng Ding[a,b], Xiao Zhang[a,b], Bo Xu[a,b], 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] N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation

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: liubin1981@nankai.edu.cn or huangdi@nankai.edu.cn

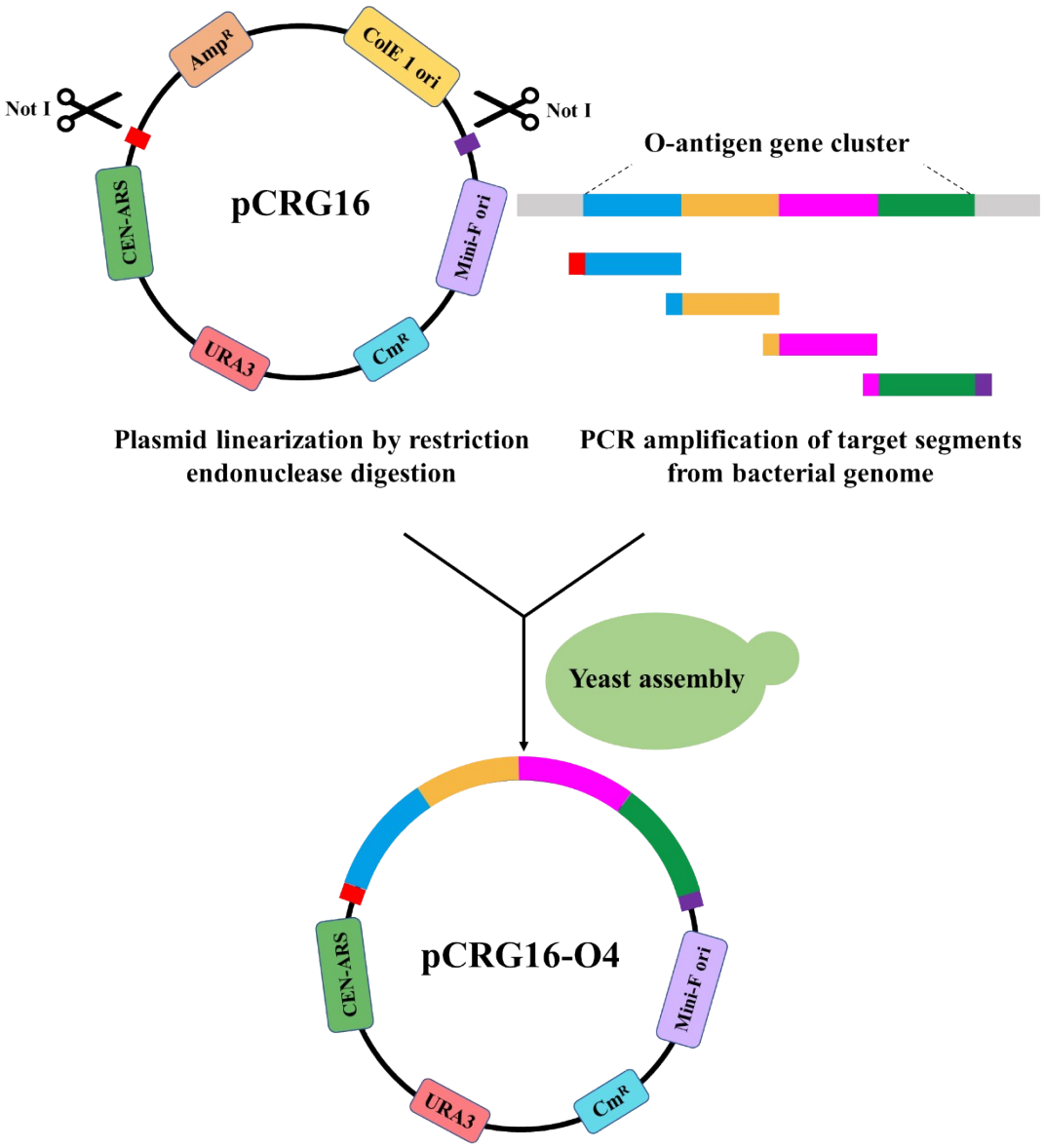

Fig. S1 Workflow for the construction of *E. coli* O4 OPS expression plasmid.

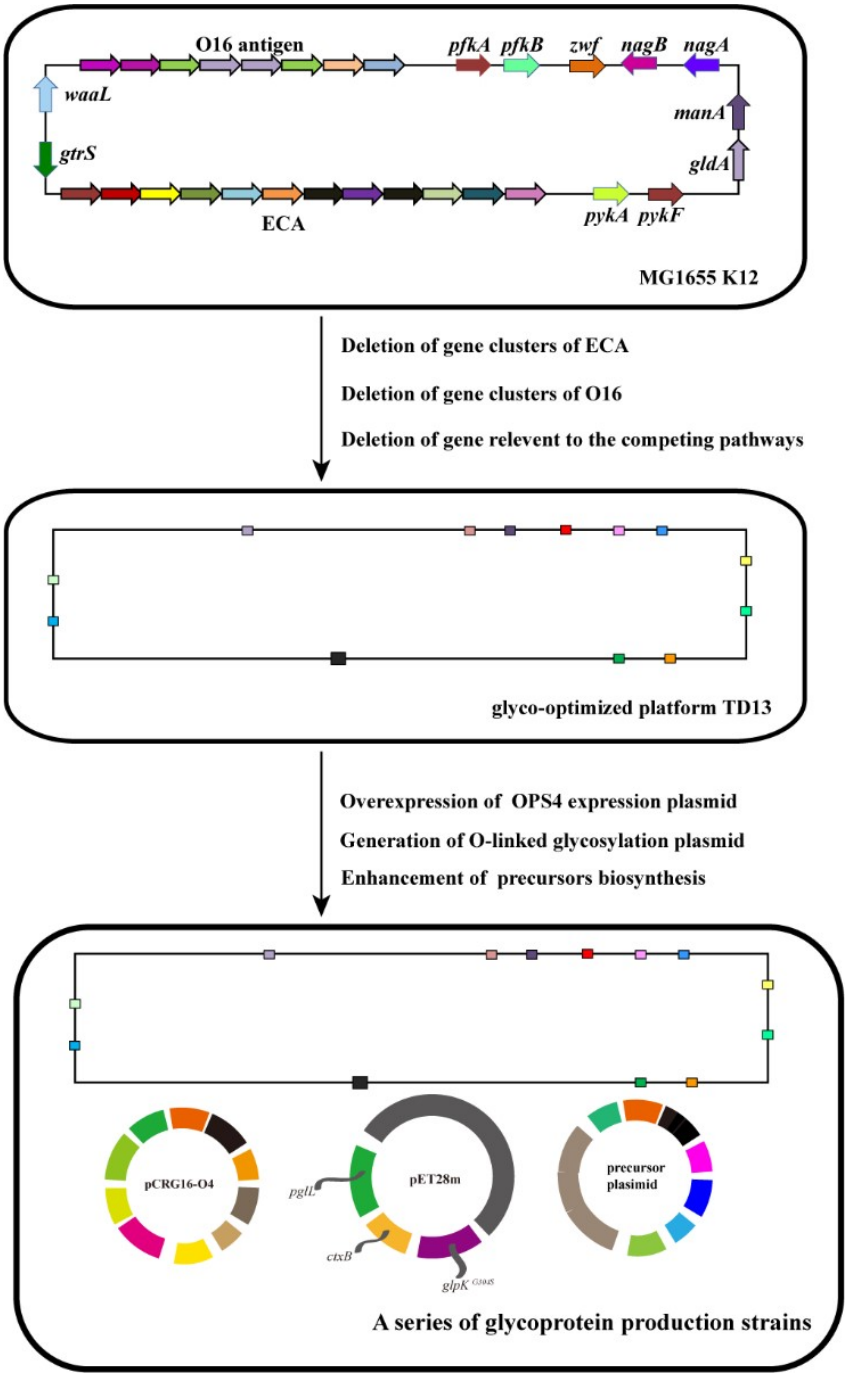

Fig. S2 The procedure for constructing glyco-optimized platform TD13 to synthesize OPS4-specific glycoprotein. Gene clusters for the biosynthesis of O16 O-antigen, ECA, in *E. coli* MG1655 were deleted in the genome. Furthermore, some genes relevant to the competing pathways and *waaL* were deleted to construct the glyco-optimized platform TD13. Gene clusters for the biosynthesis of *E. coli* O4 O-antigen, genes responsible for O-linked glycosylation and the genes relevant to the precursor biosynthesis pathways were overexpressed to generate a series of glycoprotein production strains.

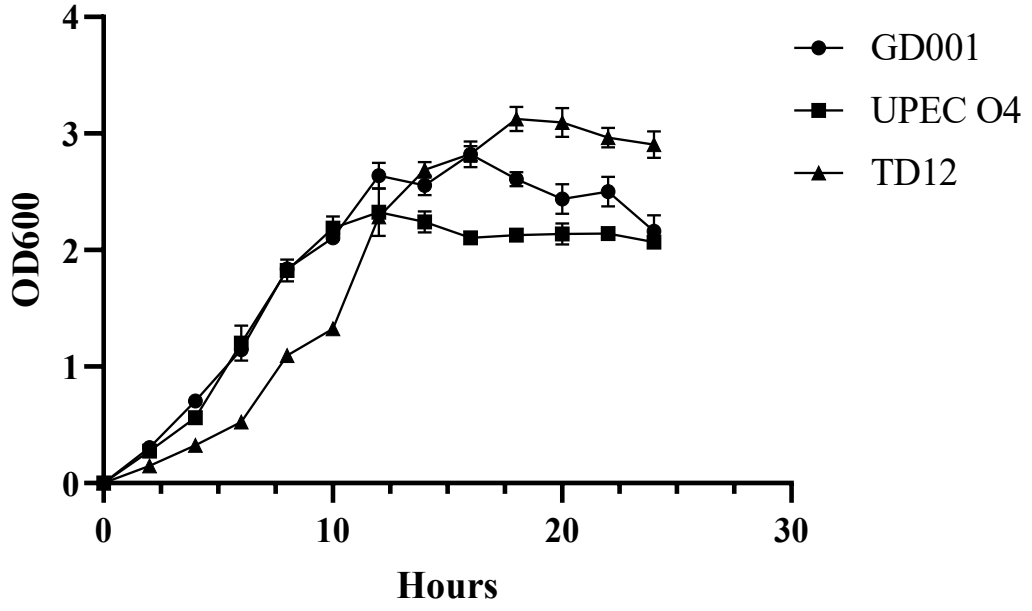

Fig. S3 The growth profiles of LPS-producing strains. Abbreviation: GD001 represents *E. coli* K-12 MG1655 harboring the OPS4 expression plasmid pCRG16-O4; TD12 represents *E. coli* K-12MG1655ΔO16::FRTΔECA*::FRTΔ*zwf* Δ*pfkA* Δ*pfkB* Δ*nagB* Δ*pykA* Δ*gldA* Δ*pykF* Δ*manA* Δ*nagA* Δ*gtrS* harboring the OPS4 expression plasmid pCRG16-O4; UPEC O4 represents uropathogenic *E. coli* O4 serotype, as a positive control in this study.

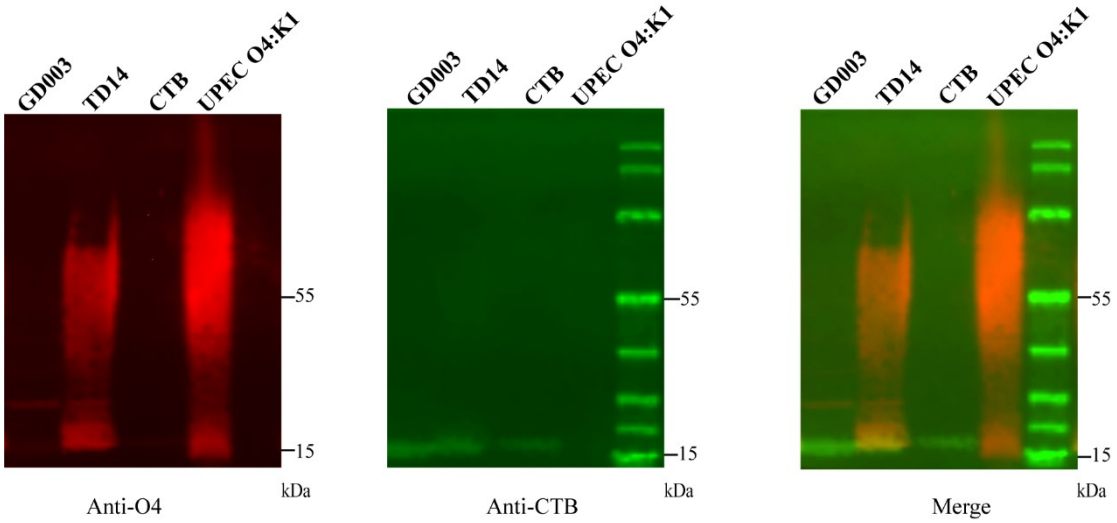

Fig. S4 Western blot analysis of recombinant glycoproteins purified from the production strain TD14 and control strain GD003. Anti-O4 channel probing with O4 OPS; Anti-CTB channel probing with CTB to verify the carrier protein; Merged images channel indicates anti-O4 and anti-CTB signals merged; Each lane was loaded with ~0.5 μg of glycoprotein based on total protein. Abbreviation: GD003 represents *E. coli* K-12 MG1655Δ*waaL* harboring the OPS4 expression plasmid pCRG16-O4 and O-linked glycosylation plasmid pet28M-*pglL-CTB-glpK*^{G304S}; TD14 represents *E. coli* K-12MG1655ΔO16::FRTΔECA*::FRTΔ*zwf* Δ*pfkA* Δ*pfkB* Δ*nagB* Δ*pykA* Δ*gldA* Δ*pykF* Δ*manA* Δ*nagA* Δ*gtrS* Δ*waaL* harboring the OPS4 expression plasmid pCRG16-O4 and O-linked glycosylation plasmid pet28M-*pglL-CTB-glpK*^{G304S}; UPEC O4:K1 represents uropathogenic *E. coli* O4 serotype, as a control in this study.

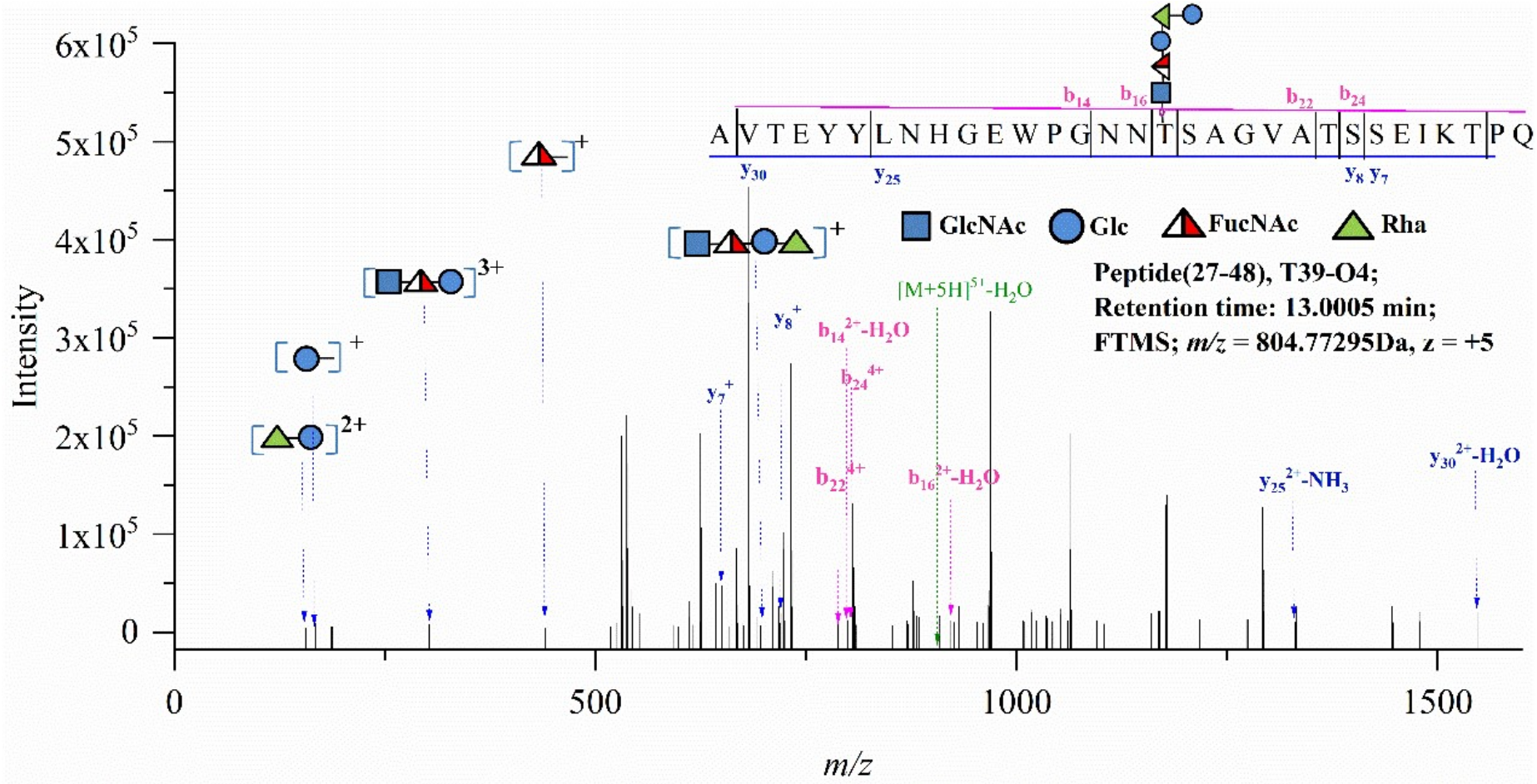

Fig. S5 Nano-LC-MS/MS analysis of purified glycoprotein generated by TD14 carrying O-linked glycosylation plasmid pET28m- *pglL-CTB-glpK*^{G304S} and the OPS4 expression plasmid.

Table S1 Strains used in this study.

Strains	Relevant characteristics	Source
DH5α	F^{-}*glnV44deoRnupGpurB20* Φ80*dlacZ*ΔM15Δ(*lacZYA-argF*)U169 *recA1endA1hsd*R17(*rk*$^{-}$, *mk*$^{+}$) *thi*-1 *gyr*A96 *relA1* λ−	Lab stock
E. coli K-12 MG1655	Wild type *E. coli* K-12; F^{-} λ− *ilvG- rfb*-50 *rph*-1	Lab stock
NEB-10β	*araD139* Δ(*ara-leu*)*7697fhu*A *lac*X74 *galK16* (f80*d*Δ(*lacZ*)M15) *galE15rec*A1 *end*A1 *relA1nupG rps*L(*Str*^{R}) Δ(*mrr-hsd*RMS *mcr*BC) *rphspoT1*	NEB
E. coli O4:K1	Clinical isolate G1625	Lab stock
GD001	*E. coli* K-12 MG1655 harboring pCRG16-O4	Lab stock
GD002	*E. coli* K-12 MG1655Δ*waaL*	Lab stock
GD003	GD002 harboring pCRG16-O4 and pet28M-*Nm*pglL-ctxB-glpK^{G304S}	Lab stock
TD11	*E. coli* K-12 MG1655ΔO16::FRT ΔECA*::FRT Δ*zwf* Δ*pfkA* Δ*pfkB* Δ*nagB* Δ*pykA* Δ*gldA* Δ*pykF* Δ*manA* Δ*nagA* Δ*gtrS*	Lab stock
TD12	TD11 harboring pCRG16-O4	This study
TD13	TD11Δ*waaL*	This study
TD14	TD13 harboring pCRG16-O4 and pet28M-*Nm*pglL-ctxB-glpK^{G304S}	This study
TD15	TD14 harboring pCDFDuet-*Ec*RmlABCD	This study

TD16	TD14 harboring pCDFDuet-RmlABCD	This study
TD17	TD14 harboring pCDFDuet-FnlABC	This study
TD18	TD14 harboring pCDFDuet-WbjBCD	This study
AGO01	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23108-FnlABC	This work
AGO02	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23108-FnlABC	This work
AGO03	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23113-FnlABC	This work
AGO04	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23119-FnlABC	This work
AGO05	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23113-FnlABC	This work
AGO06	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23119-FnlABC	This work
AGO07	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23119-FnlABC	This work
AGO08	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23108-FnlABC	This work
AGO09	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23113-FnlABC	This work
AGO10	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23108-WbjBCD	This work
AGO11	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23113-WbjBCD	This work
AGO12	TD14 carrying plasmid pCDFDuet-P23119-RmlABCD-P23119-WbjBCD	This work
AGO13	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23108-WbjBCD	This work

AGO14	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23113-WbjBCD	This work
AGO15	TD14 carrying plasmid pCDFDuet-P23113-RmlABCD-P23119-WbjBCD	This work
AGO16	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23113-WbjBCD	This work
AGO17	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23108-WbjBCD	This work
AGO18	TD14 carrying plasmid pCDFDuet-P23108-RmlABCD-P23119-WbjBCD	This work
S. cerevisiae	*S. cerevisiae* recombinational cloning strain; *MATα, ura3Δ, cyh2*^{R}	Reference[1]

Table S2 Plasmids used in this study.

Plasmids	Relevant characteristics	Source
pET28m	pet28a (+)-derived plasmid, the original T_7 region between *Bgl*II and *Xho*I site was replaced by a 1236-nucleotide fragment containing double *tac* promoter, *rrnB*_T1/T2 terminator. Kan^R	Lab stock
pCDFm	pCDFDuet (+)-derived plasmid, the original T_7 region between *Bgl*II and *Xho*I site was replaced by a 1236-nucleotide fragment containing double *tac* promoter, *rrnB*_T1/T2 terminator. Str^R	This study
pCRG16	Yeast-*E. coli* shuttle vector for recombinational cloning of large DNA fragments. *E. coli* mini-F and ColE1 replicons, yeast CEN-ARS replicon. Cm^R, Amp^R (*E. coli*); URA3, CYH2 (yeast)	Lab stock
pCRG16-O4	pCRG16 harboring the O4 gene cluster containing *wcaM* to *hisI*	This study
pET28m-*pglL-CTB*	Derivative of pet28M containing the *pglL* and *CTB*	Lab stock
pET28m-*pglL-CTB- glpK* ^{G304S}	Derivative of pet28M containing the *pglL* and *CTB* and *glpK* ^{G305S} genes. Kan^R	This study
pD	Derivative of pCDFm containing the *pgm*, *galUE*, *glmS* and *glmU* genes. Str^R	This study
pCas	*rep*A101(Ts) *pSC101 Pcas-cas*9 P_{araBAD}- *exo*, *beta* and *gamalac*I^qP_{trc}-sgRNA-*pMB1*. Kan^R	This study
pRock	p15A PlacIQ-sgRNA-ApmR J23119-sgRNA-BsaI. Tet^R	Reference [2]
pPaper	p15A PlacIQ-sgRNA-TetR J23119-sgRNA-BsaI. Cm^R	Reference [2]

Plasmids	Relevant characteristics	Source
pTarget	*pMB1aadA*. Str^R or $Spec^R$	This study
pCDF-RmlABCD	Derivative of pCDFm which carries the genes *RmlA*, *RmlB*, *RmlC* and *RmlD*. Str^R	This study
pCDF-EcRmlABCD	Derivative of pCDFm which carries the genes *EcRmlA*, *EcRmlB*, *EcRmlC* and *EcRmlD*. Str^R	This study
pCDF-FnlABC	Derivative of pCDFm which carries the genes *FnlA*, *FnlB* and *FnlC*. Str^R	This study
pCDF-WbjBCD	Derivative of pCDFm which carries the genes *WbjB, WbjC* and *WbjD.* Str^R	This study
pCDF-P23108-RmlABCD-P23108-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23108-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23108-RmlABCD-P23113-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23108-RmlABCD-P23119-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23119 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23113-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23119-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23119 promoter in MCS-2. Str^R	This study

Plasmids	Relevant characteristics	Source
pCDF-P23119-RmlABCD-P23119-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23119 promoter in MCS-2. Str^R	This study
pCDF-P23119-RmlABCD-P23108-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23119-RmlABCD-P23113-FnlABC	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *FnlA*, *FnlB* and *FnlC* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23119-RmlABCD-P23108-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23119-RmlABCD-P23113-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23119-RmlABCD-P23119-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23119 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23119 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23108-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23113-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23113-RmlABCD-P23119-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23113 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23119 promoter in MCS-2. Str^R	This study

Plasmids	Relevant characteristics	Source
pCDF-P23108-RmlABCD-P23113-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23113 promoter in MCS-2. Str^R	This study
pCDF-P23108-RmlABCD-P23108-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23108 promoter in MCS-2. Str^R	This study
pCDF-P23108-RmlABCD-P23119-WbjBCD	pCDFDuet-1 plasmid carrying the genes *RmlA*, *RmlB*, *RmlC* and *RmlD* with J23108 promoter in MCS-1, and the genes *WbjB, WbjC* and *WbjD* with J23119 promoter in MCS-2. Str^R	This study
pCDFm-pgm-galUE	Derivative of pCDFm which carries the genes *pgm* and *galUE*. Str^R	This study
pCDFm-pgm-galUE-glmS-glmU	Derivative of pCDFm which carries the genes *pgm*, *galUE, glmS* and *glmU*. Str^R	This study
pCP20	*rep*A101(Ts) *pSC101* λ repressor (ts) FLP. Amp^R and Cm^R	This study

Table S3 Oligonucleotides used in this study.

Primer	Sequence
O4-1-F	ACTTCGTATAGCATACATTATACGAAGTTATATTCGATG CGCAACATCTTTTTATGCGAAATATC
O4-1-R	GATTTCTCTGATTACAGCTCGACTA
O4-2-F	TCATACAAAACTGCAAGCCAATTAT
O4-2-R	CATCTAACCTATCAGTATTATGAGA
O4-3-F	GTTGGCTATTGCACTCAAAGAATTA
O4-3-R	TCATCGGCAATTTGCTTGAAGTAAG
O4-4-F	AAAGCTGAGTTTATCGAGAAAGTTC
O4-4-R	TGCCCGTTTTCTTTAAAACCGAAAA
O4-gtrS-F	gctgctgggcggtatggtgggtgcggggattgtgctggggcgtaacgctctgcgtaattac aacgtgaagtaatcttttcggttttaaagaaaacgggcaGAGCTTGGATGATTA TTGACACAC
O4-gtrS-R	AATTATAATTATTTTTATAGCACGTGATGAAAAGGACC GCGTGTGTCAATAATCATCCAAGCTC
O4-verify-F1	TCTCTTGCAAATGTTCATGCGATGA
O4-verify-R1	TCTGGTTAAAACTCTCAAAAAAGAA
O4-verify-F2	ACACACATTGCTATGGAATGATCCA
O4-verify-R2	CGATAATAATCACCCATATCCTCCG
O4-verify-F3	GGCAATAATGGTGATATTTTTGTCC
O4-verify-R3	TCATTTTCTTCTCGATATCGAAAGT
O4-verify-F4	AGTACAACTGGGATCTGAACTACGG
O4-verify-R4	GCTCTTCCTGATTATCCAGCGTTTC
O4-verify-F5	ATGCCCGTTTTTTTGCCGGATGCGATGC
O4-verify-R5	ATGTTAACAGAACAACAACGTCGCG
Gene-deletion construction	
PTarget F	ACTAGTATTATACCTAGGACTGAGC
PTarget R	CCTGTTATCCCTACTCGAG
gRNAR	GTAATAGATCTAAGCTTCTGC
manA-up-F	gaagcttagatctattacATCGCGGTGCCGGTATCCTGACAGG

manA-up-R	GATCAATCCCTGTTTTAATGTGGAA
manA-down-F	TTCCACATTAAAACAGGGATTGATCCATCTCTTGCTAAG CTTATT
manA-down-R	tcgagtagggataacaggACTGGAATCACCGCTGTAACCAATG
manA-gRNA	cctaggtataatactagtCTATGCCTGGGGCAGCAAAAgttttagagctag aaatagca
wecB-up-F	gaagcttagatctattacTCAATGACAGCGTGAAGCTTATTGC
wecB-up-R	CGATTCTCTTCGAATAAGCGGCGAG
wecB-down-F	CTCGCCGCTTATTCGAAGAGAATCGGCGACCATTTCTGT TATCGG
wecB-down-R	tcgagtagggataacaggTGCTGCGGGAAAGTGAGATCCGGAC
wecB-gRNA	cctaggtataatactagtCAACATAATGCAGCCAGGACgttttagagctag aaatagca
waaL-up-F	gaagcttagatctattacGTAGAACGCACCCGCGAACTGTTTG
waaL-up-R	TCCAAAACTGCTTGACTGTTGAGGT
waaL-down -F	ACCTCAACAGTCAAGCAGTTTTGGAAGTTTAAGAAGTG AGTTAAAACTCA
waaL-down -R	tcgagtagggataacaggGAATTTATTGGAGATACAAGAATAG
waaL-gRNA	cctaggtataatactagtTAGGCTTGCTTGATTTAATTgttttagagctaga aatagca
glpK^{G3054}-F	GTATACATGACTGAAAAAAAATATATCGTTG
glpK^{G3054}-R	GATGGCGTGGGAAGAACACGACGAA
pCDF-EcRmlABCD-F1	CCTGCATTGTGAAAATACTTGTTACTGGTGGCG
pCDF-EcRmlABCD-R1	GGATCCTTACTGGCGGCCCTCATAGTTCTCT
pCDF-EcRmlABCD-F2	GGATCCAAGGAGATATACATGAATATCCTCCTTTTTGG C
pCDFm-EcRmlABCD-R2	GAGCTCTTAAATTGCCGTAGTCGTAAATA
pCDFm-EcRmlABCD-F3	GAGCTCAAGGAGATATACATGAAAACGCGTAAAG
pCDFm-EcRmlABCD-R3	CTGCAGTTAATAACCTTTAATCATTTTTAGC
pCDFm-EcRmlABCD-F4	CTGCAGATGAACGTAATTAAAACTGAAATTC
pCDFm-EcRmlABCD-R4	AAGCTT TTATATTGTAGTTGGATCATTCCAT
pCDFm-EcRmlABCD-VF1	TTGACAGCTAGCTCAGTCCTAGGTA

-EcRmlABCD-VR1	GGCGCTAAAGCGCGTTGCAGTTCGT
pCDFm-EcRmlABCD-VF2	TGAACGCGGGCGCGTATACCGCGGT
-EcRmlABCD-VR2	CTGCCAGGCATCAAATAAAACGAGC
pCDFm-RmlABCD-F1	ATGAAACGCAAGGGCATCATCCTCG
pCDFm-RmlABCD-R1	TCAGTACACGGTCTCGGTCAGCAGG
pCDFm-RmlABCD-F2	ATTCGCGAAATTCTGATTATTAGCA
pCDFm-RmlABCD-R2	CCGGTATCCAGCCACGCATAGCCGC
pCDFm-RmlABCD-F3	ATGAAAGCGACCCGCCTGGCAATTCCCG
pCDFm-RmlABCD-R3	TCAGGGGAAGCAGTCGGCGTCGGCG
pCDFm-RmlABCD-F4	ATGAACCGGATCCTTCTCCTCGGCG
pCDFm-RmlABCD-R4	TCATAGTGGTCCCTGCTCGCTCA
pCDFm-RmlABCD-VF1	CTAGCTCAGTCCTAGGTATAATAAAG
pCDFm-RmlABCD-VR1	CGGTCAGCAGGCGTTTCAGATACTG
pCDFm-RmlABCD-VF2	AAAACCTATGCGGGCAACCGCAACG
pCDFm-RmlABCD-VR2	TTACGCATATTTGCCCACC
pCDFm-FnlABC-F1	ATGTTTAATGGTAAAATATTGTTAA
pCDFm-FnlABC-R1	TTATGAATCCAACTCATAATCTTCA
pCDFm-FnlABC-F2	ATGAGTTGGATTCATAATATGAAAA
pCDFm-FnlABC-R2	TCACGATAAAACTCTCGCTATAGTA
pCDFm-FnlABC-F3	ATGTCGGTTGTTGGGACTCGTCCAG
pCDFm-FnlABC-R3	TTAGTATTGCTTCCAGACAACCCGTT
pCDFm-FnlABC-VF1	GAGATATACATGTTTAACGGCAAAA
pCDFm-FnlABC-VR1	GGCGCTTTCTGCACAAAAATATCGC
pCDFm-FnlABC-VF2	TACTGAAACTGCCGTTTATTCGCGC
pCDFm-FnlABC-VR2	GTATACGCAAACTGTTCCGGGCTCA
pCDFm-WbjBCD-F1	CATATGATGGATAAGAACTCTGTTCTGTTAAT
pCDFm-WbjBCD-R1	GATATCTTTTCCTCTGGACTGGCACTCTCGA
pCDFm-WbjBCD-F2	GATATCATGAAAGTTCTTGTAACTGGCGCGA
pCDFm-WbjBCD-R2	GGTACC TTACGCACCTTCGCCTACTGAACAA
pCDFm-WbjBCD-F3	GGTACCATGCAGAAGCTAAAAGTCGTTACGG
pCDFm-WbjBCD-R3	CTCGAGTTAATATTTTTTCCAGACAGTTCGC

pCDFm-WbjBCD-VF1	ATGAGCTGTTGACAATTAATCATC
pCDFm-WbjBCD-VR1	ATCTCCTTAGTAGGCTATTCTTCCG
pCDFm-WbjBCD-VF2	TACCTATGCGTGCAGCGTGGGTGAA
pCDFm-WbjBCD-VR2	GTGCGCATCACATAATCGCGATAGC

Lowercase letters represent homologous sequences to the genome, RBS sequences, restriction site sequences. Red letters represent N20 sequences of each gene.

Table S4 *E*-factors and process complexity determined for the production process of glycoprotein.

Manufacturing processes	*E*-factor[a] [kg kg^{-1}]	No. of steps
Bioconjugation methodology	1600~2000[b]	3-4
Pharmaceuticals (antibiotics, drugs, vaccines)	25−＞100[c]	2+

[a] The *E*-factor (Environmental impact factor) was proposed for assessing the environmental impact of manufacturing processes[4].

$$\frac{\sum m\ (\text{Input materials}) - m(\text{Product})}{m(\text{Product})}$$

[b] Materials (glucose, glycerol and other raw materials) usage in green manufacture were described. Furthermore, losses from the purification process are also taken into account.

[c] The *E* factor in the pharmaceutical industry was typically calculated as 25−＞100 kg per kg[5].

The detailed calculation process of *E* factor is displayed as followed.

The American Chemical Society (ACS) Green Chemistry Institute Pharmaceutical Roundtable (GCIPR) group has an initial attempt to establish a systematic environmental assessment of therapeutic biologics[6].

The *E* factor in the pharmaceutical industry was typically calculated as 25 to over 100 kg per kg. However, traditional chemical methods for the production of glycoproteins are usually complex. Multiple steps are required whereby the O-antigen must be purified from the pathogen of interest, detoxified and subject to chemical activation. In parallel, the protein must also be purified and chemically activated before protein and glycan can be conjugated. Following conjugation, further rounds of purification are necessary before glycoprotein can be administered. Therefore, it takes at least 4 steps rather than 2 steps as reported to synthesize such glycoproteins using chemical and/or chemoenzymatic synthetic techniques.

Analysis of the bioconjugation route, the *E* factor was estimated as 1600-2000 kg per kg. A primary cause of the high *E* factors within the bioconjugation methodology is the rather low glycoprotein titer and the correspondingly large amount of input materials. In fact, the strategies for the biosynthesis of the glycoproteins offer a

promising alternative to prepare the therapeutic glycoprotein as a vaccine against UPEC in the microbial cell factory. However, this therapeutic glycoprotein is still in its initial stage, while proof-of-concept reactions have been demonstrated, further stepwise improvements are required to enhance the glycosylation efficiency (or site occupancy) for industrial production. For *E* factor calculation, it takes the chemical yield into account and includes reagents, solvent losses, all process aids and, in principle, even fuel (although this is often difficult to quantify). However, the glucose and glycerol and many other chemicals in this medium are not completely metabolized during the fermentation process. After 48 h fermentation, the content of the residual glucose and glycerol in the culture was at least 10 g/L and 10 g/L. Hence, the *E* factor in the bioconjugation route ranged from 1600 to 2000.

Despite the high *E* factor, such recombinant *E. coli* can direct the biosynthesis of glycoprotein in a "living glyco-factory" without the use of purified enzymes and protective groups usually applied by traditional chemical technologies.

Furthermore, the production of glycoprotein in glycoengineering *E. coli* can reduce the risk of viral contamination, which is superior to current chemical coupling methods. Moreover, the simple, cost-efficient and rapid growth culturing of the *E. coli* glycoengineering strain expands this method to a wider application of laboratories than existing chemical/chemoenzymatic production technologies.

The 12 Principles of Green Chemistry were developed by Paul Anastas and John Warner as a way to "help reduce or eliminate the use or generation of hazardous substances in the design, manufacture and application of chemical products"[7]. A traffic light assessment of the oligonucleotide process against these principles is presented in Table S5, highlighting several areas for improvement.

Table S5 Assessment of glycoprotein manufacturing process against 12 principles of green chemistry.

Principle	Assessment for bioconjugation method (initial stage)	Assessment for chemical methods
Prevention of Waste	A rather low titer which inevitably generates many wastes (mainly wastewater and carbon mix)	The generated waste are slightly low.
Atom Economy	Atom economy for the starting materials is slightly poor	Atom economy for the starting materials is high
Less Hazardous Chemical Syntheses	No or less hazardous reagents and solvents are involved	Some hazardous reagents and solvents are involved
Designing Safer Chemicals	Glycoprotein generally as therapeutic drugs	Glycoprotein generally as therapeutic drugs
Safer Solvents and Auxiliaries	Only the safety fermentation medium is used	Some solvents are used in the detoxified step
Design for Energy Efficiency	Medium Energy Efficiency	Medium Energy Efficiency
Use of Renewable Feedstocks	Most reagents and materials are renewable and green feedstocks	Some reagents and materials are nonrenewable feedstocks
Reduce Derivatives	No Derivatives generation	Protecting groups are used
Catalysis	Biocatalysis process in vivo using the OTase	Chemical conjugation process used
Design for Degradation	Can be degraded in *vivo*	Can be degraded
Real-time Analysis for Pollution Prevention	Analysis methods need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances	Analysis methods need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances
Inherently Safer Chemistry for Accident Prevention	Inherently safer chemistry for accident prevention	Purification and detoxification of polysaccharide from the pathogenic bacteria

References

1. C. K. Raymond, E. H. Sims and M. V. Olson, *Genome Res.*, 2002, **12**, 190-197.
2. J. Wang, X. Sui, Y. Ding, Y. Fu, X. Feng, M. Liu, Y. Zhang, M. Xian and G. Zhao, *Nucleic Acids Res.*, 2021, **49**, e12.
3. K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci. U.S. A.*, 2000, **97**, 6640-6645.
4. R. A. Sheldon, *Green Chem.*, 2017, **19**, 18-43.
5. F. Roschangar, R. A. Sheldon and C. H. Senanayake, *Green Chem.*, 2015, **17**, 752-768.
6. S. V. Ho, in *Green Chemistry in the Pharmaceutical Industry*, ed. P.J. Dunn, A.S. Wells and M.T. Williams, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 1st edition, 2010, Chapter 15: Environmental Considerations in Biologics Manufacture, pp. 311-331.
7. P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, Oxford, 1998.