Supplementary Information

Engineering orthogonal human *O*-linked glycoprotein biosynthesis in bacteria

Aravind Natarajan,[1] Thapakorn Jaroentomeechai,[2] Marielisa Cabrera-Sánchez,[1] Jody C. Mohammed,[2] Emily C. Cox,[3] Olivia Young,[2] Asif Shajahan,[4] Michael Vilkhovoy,[2] Sandra Vadhin,[2] Jeffrey D. Varner,[2] Parastoo Azadi,[4] and Matthew P. DeLisa[1,2,3]*

[1]Department of Microbiology, Cornell University, Ithaca, NY 14853 USA
[2]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, 120 Olin Hall, Ithaca, NY 14853 USA
[3]Biomedical and Biological Sciences, Cornell University College of Veterinary Medicine, Ithaca, New York, 14853 USA
[4]Complex Carbohydrate Research Center, The University of Georgia, Athens, GA 30602 USA.

Supplementary Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant genotype	Reference
E. coli strains		
DH5α	F⁻ φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17* (r_k^-, m_k^+) *gal-phoA supE44* λ⁻ *thi-1 gyrA96 relA1*	Laboratory stock
NEB 10-beta	*araD139* Δ(*ara-leu*)7697 *fhuA lacX74 galK* (f80 Δ(*lacZ*)M15) *mcrA galU recA1 endA1 nupG rpsL* (Str^R) Δ(*mrr-hsdRMS-mcrBC*)	New England Biolabs
BL21(DE3)	F⁻ *ompT hsd*S_B (r_B^-, m_B^-) *gal dcm* (DE3)	Laboratory stock
W3110	F⁻ λ⁻ *rph-1* IN(*rrnD-rrnE*)1	Laboratory stock
CLM24	W3110 Δ*waaL*	[1]
CLM25	CLM24 Δ*wecA*	This study
MC4100	F⁻ *araD139* Δ(*argF-lac*)*169* λ⁻ *e14⁻ flhD5301* Δ(*fruK-yeiR*)*725(fruA25) relA1 rpsL150*(Str^R) *rbsR22* Δ*(fimB-fimE)632(*::IS*1*) *deoC1*	Laboratory stock
MCΔw	MC4100 Δ*wecA*	This study
MCΔΔw	MC4100 Δ*wecA* Δ*waaL*	This study
MCΔΔw-*neu*$_{O\text{-}PS}$	MC4100 Δ*wecA* Δ*waaL* with *neuDBAC* genes at O-PS site	This study
MCΔΔwΔn-*neu*$_{O\text{-}PS}$	MCΔΔw-*neu*$_{O\text{-}PS}$ Δ*nanA*	This study
ZLKA	DH1 *lacZ lacA nanKETA*	[2]
Plasmids		
pMW07	Yeast-based recombineering plasmid with yeast origin of replication and URA3 selection marker; Cm^R	[3]
pMW08	Modified plasmid pMW07 with the yeast origin of replication and URA3 selection marker removed; Cm^R	This study
pOG-Tn	Genes encoding *C. jejuni* Gne and *A. baumanii* PglC cloned in plasmid pMW08; Cm^R	This study
pOG-Tn-*HsC1*GalT1	Gene encoding glycoprotein-*N*-acetylgalactosamine 3-β-galactosyltransferase 1 from *H. sapiens* cloned without the first 29 amino acids in plasmid pOG-Tn; Cm^R	This study
pOG-Tn-*DmC1*GalT2	Gene encoding glycoprotein-*N*-acetylgalactosamine 3-β-galactosyltransferase A, isoform B from *D. melanogaster* cloned without the first 50 amino acids in plasmid pOG-Tn; Cm^R	This study
pOG-Tn-BiGalHexNAcP	Gene encoding D-galactosyl-β1-3-*N*-acetyl-D-hexosamine phosphorylase from *B. longum* subspecies *infantis* cloned in plasmid pOG-Tn; Cm^R	This study
pOG-Tn-*Cj*CgtB	Gene encoding S42 mutant of β1-3-galactosyltransferase from *C. jejuni* cloned in plasmid pOG-Tn; Cm^R	This study
pOG-Tn-*Ec*WbnJ	Gene encoding 1,3-α-*N*-acetylgalactosamine-diphospho-undecaprenol β-1,3-galactosyltransferase from *E. coli* O86 cloned in plasmid pOG-Tn; Cm^R	This study
pOG-Tn-*Ec*WbwC	Gene encoding *N*-acetylgalactosamine-diphospho-undecaprenol β1,3-galactosyltransferases from *E. coli* O104 cloned in plasmid pOG-Tn; Cm^R	This study
pOG-TΔ*gne*	Same as pOG-Tn-*Ec*WbwC but lacking *Cj*Gne epimerase; Cm^R	This study
pOG-Tn-*Ng*PglO	Genes encoding *C. jejuni* Gne*, A. baumanii* PglC, and *N. gonorrhea* PglO in plasmid pMW07; Cm^R	This study
pOG-Tn-*Nm*PglL	Genes encoding *C. jejuni* Gne*, A. baumanii* PglC, and *N. meningitidis* PglL in plasmid pMW07; Cm^R	This study
pOG-T	Genes encoding *C. jejuni* Gne, *A. baumanii* PglC, and *E. coli* O104 WbwC in plasmid pMW07; Cm^R	This study
pOG-T-*Ng*PglO	Gene encoding PglO from *N. gonorrhea* in plasmid pOG-T; Cm^R	This study
pOG-T-*Nm*PglL	Gene encoding PglL from *N. meningitidis* in plasmid pOG-T; Cm^R	This study
pCP20	Plasmid encoding the FLP recombinase; temperature-sensitive replication and thermal induction of FLP synthesis; Amp^R, Cm^R	[4]
pKD46	Plasmid encoding the λ-red recombinase; Amp^R	[5]
pE-FLP	Plasmid encoding the FLP recombinase; Amp^R	[6]
pRecO-PS	Shuttle vector for integration into the O-PS locus of *E. coli* K12 strains; Amp^R, Kan^R	[7]

pRecO-PS*neuDBAC*	*E. coli* K1 *neuDBAC* genes cloned into pRecO-PS; Amp^R, Kan^R	This study
pMLBy	pMLBAD vector with yeast origin of replication and URA3 selection marker; Tmp^R	Laboratory stock
pConNeuDBAC	Plasmid encoding the *E. coli* K1 *neuDBAC* genes in plasmid pMLBy with the *araC* gene and pBAD promoter replaced with the J23100 constitutive promoter from the Anderson library; Tmp^R	This study
pEXT20	IPTG-inducible expression vector; Ap^R	[8]
pEXT-spDsbA-MBP^{MOOR}	Gene encoding *E. coli* maltose-binding protein (MBP) with an *E. coli* DsbA signal peptide in place of its native signal peptide and a C-terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-$MBP^{MOORmut}$	Gene encoding *E. coli* MBP with an *E. coli* DsbA signal peptide in place of the native signal peptide and a C-terminal fusion bearing the 25-residue MOOR sequence with a Ser-to-Gly mutation in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-MBP^{MOOR}-*Ec*WbwA	Same as pEXT-spDsbA-MBP^{MOOR} but with sialyltransferase *Ec*WbwA cloned in tandem; Ap^R	This study
pEXT-spDsbA-MBP^{MOOR}-*P*spST6	Same as pEXT-spDsbA-MBP^{MOOR} but with sialyltransferase *P*spST6 cloned in tandem; Ap^R	This study
pEXT-spDsbA-YebF-MBP^{MOOR}	Gene encoding *E. coli* YebF with an *E. coli* DsbA signal peptide in place of its native signal peptide and a C terminal fusion with MBP and the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-GST^{MOOR}	Gene encoding *E. coli* glutathione-S-transferase (GST) with an *E. coli* DsbA signal peptide and a C terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-scFv13-$R4^{MOOR}$	Gene encoding single-chain Fv (scFv) antibody fragment specific for *E. coli* β-galactosidase with an *E. coli* DsbA signal peptide and a C terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-sfGFP^{MOOR}	Gene encoding superfolder green fluorescent protein (sfGFP) with an *E. coli* DsbA signal peptide and a C terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-sfGFP$^{Q157\text{-}MOOR}$	Gene encoding sfGFP with an *E. coli* DsbA signal peptide and an 25-residue MOOR sequence internally grafted at position Q157 in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-CRM197^{MOOR}	Gene encoding cross-reacting material 197 (CRM197) with an *E. coli* DsbA signal peptide and a C terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-PD^{MOOR}	Gene encoding the *Haemophilus influenzae* Protein D (PD) with an *E. coli* DsbA signal peptide and a C terminal fusion bearing the 25-residue MOOR sequence in plasmid pEXT20; Ap^R	This study
pEXT-spDsbA-MBP^{EPO}	Same as pEXT-spDsbA-MBP^{MOOR} but with 8-residue motif derived from human erythropoietin in place of the MOOR core sequence (WPAAASAP); Ap^R	This study
pEXT-spDsbA-MBP^{GPC}	Same as pEXT-spDsbA-MBP^{MOOR} but with 8-residue motif derived from human glycophorin C in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{SAP}	Same as pEXT-spDsbA-MBP^{MOOR} but with 9-residue synthetic "SAP" motif (SAPSAPSAP) in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{MUC1_8}	Same as pEXT-spDsbA-MBP^{MOOR} but with 8-residue motif derived from human MUC1 in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{MUC1_12}	Same as pEXT-spDsbA-MBP^{MOOR} but with 12-residue motif derived from human MUC1 in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{MUC1_16}	Same as pEXT-spDsbA-MBP^{MOOR} but with 16-residue motif derived from human MUC1 in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{MUC1_20}	Same as pEXT-spDsbA-MBP^{MOOR} but with 20-residue motif derived from human MUC1 in place of the MOOR core sequence; Ap^R	This study
pEXT-spDsbA-MBP^{MUC1_24}	Same as pEXT-spDsbA-MBP^{MOOR} but with 24-residue motif derived from human MUC1 in place of the MOOR core sequence; Ap^R	This study

pEXT-spDsbA-MBP^{MUC1_41}	Same as pEXT-spDsbA-MBP^{MOOR} but with 41-residue motif derived from human MUC1 in place of the entire MOOR; Ap^R	This study
pVITRO1-Trastuzumab-IgG1/κ	Genes encoding HER2/neu receptor-specific humanized IgG1/κ antibody isotype cloned in plasmid pVITRO1; Hyg^R	Addgene plasmid #61883
pVITRO1-5E5-IgG1/κ	Genes encoding Tn-MUC1-specific chimeric IgG1/κ antibody isotype cloned in plasmid pVITRO1; Hyg^R	This study
pJL1-MBP^{MOOR}	Gene encoding MBP^{MOOR} in plasmid pJL1; Kan^R	This study
pJL1-MBP^{MOORmut}	Gene encoding MBP^{MOORmut} in plasmid pJL1; Kan^R	This study

References

[1] Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, et al. Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. Proc Natl Acad Sci U S A. 2005;102:3016-21.

[2] Fierfort N, Samain E. Genetic engineering of *Escherichia coli* for the economical production of sialylated oligosaccharides. J Biotechnol. 2008;134:261-5.

[3] Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, et al. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nat Chem Biol. 2012;8:434-6.

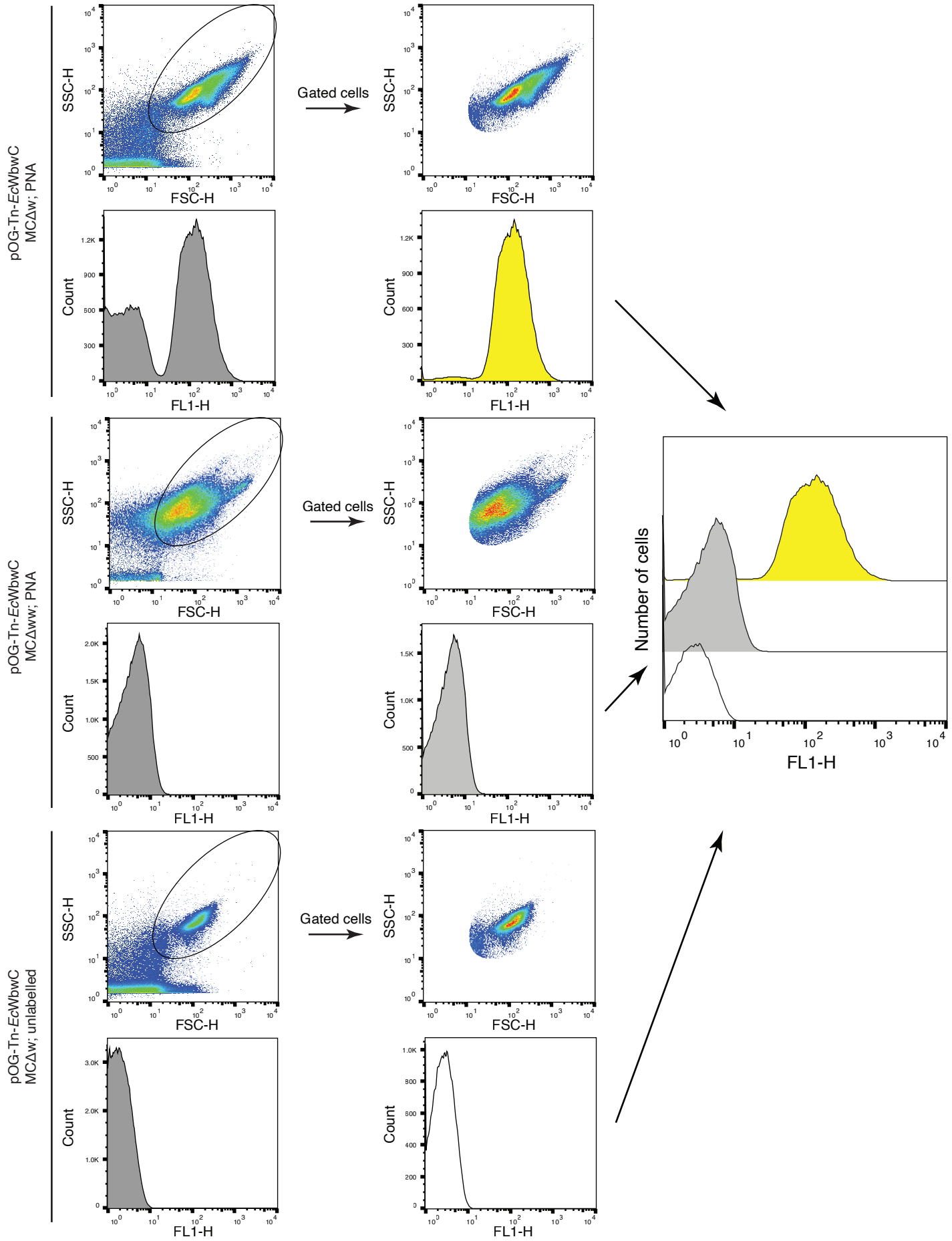
[4] Cherepanov PP, Wackernagel W. Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene. 1995;158:9-14.

[5] Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97:6640-5.

[6] St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, Shearwin KE. One-step cloning and chromosomal integration of DNA. ACS Synth Biol. 2013;2:537-41.

[7] Yates LE, Natarajan A, Li M, Hale ME, Mills DC, DeLisa MP. Glyco-recoded *Escherichia coli*: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters. Metab Eng. 2019;53:59-68.

[8] Dykxhoorn DM, St Pierre R, Linn T. A set of compatible *tac* promoter expression vectors. Gene. 1996;177:133-6.



Supplementary Figure 1. FACS gating strategy. For all flow cytometric screening, cells were analyzed using a FACSCalibur flow cytometer (BD Biosciences), and at least 100,000 total events were recorded. The events from the unlabelled MCΔw control sample were analyzed using FlowJo 10.5, and gated based on forward scatter (FSC) and side scatter (SSC) to represent the *E. coli* cell population, minimizing artifacts from debris. This same gate was then applied to all samples, followed by calculation of the median fluorescent intensity.