# Title

# Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles


## Authors

Jasmine M. Hershewe[a,b,c,1], Katherine F. Warfel[a,b,c,1], Shaelyn M. Iyer[a], Justin A. Peruzzi[a,b,c], Claretta J. Sullivan[d], Eric W. Roth[e], Matthew P. DeLisa[f,g,h], Neha P. Kamat[b,c,i], and Michael C. Jewett[a,b,c,j,k 2]

## Affiliations

[a]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, IL 60208, USA
[b]Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, USA
[c]Center for Synthetic Biology, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, IL 60208, USA
[d]Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
[e]Northwestern University Atomic and Nanoscale Characterization and Experimentation (NU*ANCE*) Center, 2145 Sheridan Road, Tech Institute A/B Wing A173, Evanston, IL 60208, USA
[f]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, USA
[g]Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA
[h]Biomedical and Biological Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA
[i]Department of Biomedical Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E310, Evanston, IL 60208, USA
[j]Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North Saint Clair Street, Suite 1200, Chicago, IL 60611, USA
[k]Simpson Querrey Institute, Northwestern University, 303 East Superior Street, Suite 11-131, Chicago, IL 60611, USA

[1]Authors contributed equally
[2]To whom correspondence should be addressed
*Michael C. Jewett
**Phone:** 1 847-497-5007
**Email:** m-jewett@northwestern.edu

**Abstract**

Cell-free gene expression (CFE) systems from crude cellular extracts have attracted much attention for biomanufacturing and synthetic biology. However, activating membrane-dependent functionality of cell-derived vesicles in bacterial CFE systems has been limited. Here, we address this limitation by characterizing native membrane vesicles in *Escherichia coli*-based CFE extracts and describing methods to enrich vesicles with heterologous, membrane-bound machinery. As a model, we focus on bacterial glycoengineering. We first use multiple, orthogonal techniques to characterize vesicles and show how extract processing methods can be used to increase concentrations of membrane vesicles in CFE systems. Then, we show that extracts enriched in vesicle number also display enhanced concentrations of heterologous membrane protein cargo. Finally, we apply our methods to enrich membrane-bound oligosaccharyltransferases and lipid-linked oligosaccharides for improving cell-free *N*-linked and *O*-linked glycoprotein synthesis. We anticipate that these methods will facilitate on-demand glycoprotein production and enable new CFE systems with membrane-associated activities.

## Manuscript Text

## Introduction

Cell-free gene expression (CFE) systems activate transcription and translation using crude cellular extracts instead of living, intact cells.[1] In recent years, these systems have matured from widely used tools in molecular biology to platforms for biomanufacturing and synthetic biology.[1–4] Among CFE systems, *Escherichia coli*-based methods have been used the most[1,5–10]. A body of work dedicated to optimization of extract preparation and reaction conditions has simplified, expedited, and improved the cost and performance of *E. coli* CFE systems[1,11–13]. Optimized *E. coli*-based CFE reactions: (i) quickly synthesize grams of protein per liter in batch reactions[14–18], (ii) are scalable from the nL to 100 L scale[19,20], and (iii) can be freeze-dried for months of shelf-stability[1,12,21–27]. The ability to readily store, distribute, and activate freeze-dried cell-free systems by simply adding water has opened new opportunities for point-of-use biosensing[28–33], portable therapeutic and vaccine production[23,24,34], and educational kits[1,35–37].Thus, efforts to improve the efficiency and expand capabilities of engineered CFE systems could have impacts across many disciplines.

In a growing number of contexts, CFE extracts have been tailored to new applications by pre-enriching soluble, heterologous components *in vivo* prior to cell lysis, avoiding the need to separately purify and add these components. Examples where cells for CFE systems have been engineered prior to generating extract include incorporation of site-specific non-canonical amino acids into proteins[15,38,39], biosensing of analytes[32,40,41], and assembly of metabolic pathways for production of small molecules[9,42–44]. Heterologous components can also be directly added to the CFE system to achieve a desired function. For example, membrane augmented cell-free systems achieved through the use of nanodiscs, synthetic phospholipid structures, purified microsomes, and purified vesicles have all enabled the use of membrane components in CFE systems[45–51].

Enabling membrane-associated functions in CFE systems is important for numerous biological functions. Indeed, lipid membranes play pivotal roles across all domains of life, with ~20-30% of genes encoding for membrane proteins and many essential processes taking place on and across membranes[52,53]. For example, membranes are required for molecular transport, immunological defense, energy regeneration, and post-translational protein modification (e.g., glycosylation).

Despite the absence of intact cellular membranes, membrane structures are present in crude extract-based CFE systems. They form upon fragmentation and rearrangement of cell membranes during cell lysis and extract preparation and have been studied and characterized for decades[51,54–57]. For example, oxidative phosphorylation and protein translocation were originally studied from purified vesicles prepared from *E. coli* cell extracts.[58] In *E. coli* CFE systems, inverted membrane vesicles harboring electron transport chain machinery activate oxidative phosphorylation and ATP regeneration[54,59]. Analogously, in eukaryotic-derived crude extract-based CFE systems, endoplasmic reticulum (ER)-derived microsomes enhance functionality, enabling the synthesis of membrane proteins and proteins with disulfide bonds, among others[5,27,60–65].

While there are limited examples of cell-derived membrane-incorporated components to enhance bacterial CFE systems, this area of research has remained largely under-studied (with membrane augmented systems like nanodiscs being used most frequently). Yet, enriching native membrane-bound components in CFE systems, especially with heterologously expressed cargo, is poised to enable compelling applications. For example, protein glycosylation, which can profoundly impact folding, stability, and activity of proteins and therapeutics[66–69], is mediated by membrane-bound components. Introduction of cell-derived vesicles with machinery required for glycosylation could enable cell-free biomanufacturing of protein therapeutics and conjugate vaccines.

Along these lines, we recently described cell-free glycoprotein synthesis (CFGpS), a platform for one-pot biomanufacturing of defined glycoproteins in extracts enriched with heterologous, membrane-bound glycosylation machinery[34]. To date, CFGpS has been used to produce model glycoproteins, human glycoproteins, and protective conjugate vaccines[24,34,70–72]. Unfortunately, the existing CFGpS system based on S30 extracts (i.e., cell extracts that result from a 30,000 *x g* clarification spin) is limited by glycosylation efficiency, only producing ~10-20 μg/mL of glycoprotein in batch.[2,34] Characterizing and enriching cell-derived vesicles comprising membrane-bound glycosylation components offers one strategy to address this limitation, and perhaps make possible a variety of applications involving membrane-bound biology.

Here, we set out to develop methods that enhance membrane-dependent functionality of cell-derived vesicles in bacterial CFE systems, with a focus on bacterial glycoengineering. First, we characterize size distributions and concentrations of native membrane vesicles in extracts, providing a benchmark for analysis and engineering of CFE systems. To do so, we apply canonical strategies (e.g., TEM), but also apply simple and expedited characterization workflows that rely on techniques such as light scattering to directly analyze vesicles in extracts without the need for lengthy protocols. Second, we investigate the impacts of upstream extract processing steps on vesicle profiles, revealing simple handles to modulate vesicle concentration in extracts. Third, we use cell-derived membrane vesicles to enrich a variety of heterologous, membrane-bound proteins and substrates in extracts without the use of synthetically-derived membranes. Finally, we apply our findings to improve glycoprotein yields in our existing asparagine-linked (*N*-linked) CFGpS system and a new membrane-dependent CFGpS system based on serine/threonine-linked (*O*-linked) glycosylation. By applying our optimized methods to increase concentrations of vesicle-bound glycosylation machinery, we shorten the time associated with extract preparation, increase glycosylation efficiencies, and enhance glycoprotein titers by up to ~170%. Importantly, we go on to show that improvements in

glycoprotein titers are generalizable to multiple glycoproteins without the need to re-optimize conditions.

## Results

In this study, we aimed to characterize and engineer membrane vesicles (which form upon fragmentation of cell membranes during cell lysis) in *E. coli* CFE extracts (**Fig. 1**). Then, we used this knowledge to control enrichment of membrane-bound components for enhancing defined function, including improving glycoprotein synthesis (**Fig. 1**). To achieve these goals, we (i) use nanocharacterization techniques to determine the sizes and quantities of membrane vesicles in *E. coli* extracts; (ii) determine how extract processing can control the enrichment of vesicles in extracts; (iii) enrich several heterologous, membrane-bound components in extracts via vesicles; and (iv) demonstrate that increasing enrichment of membrane-bound components significantly improves cell-free glycoprotein synthesis systems for *N*- and *O*-linked glycosylation.

### *Characterization of membrane vesicles in CFE extracts*

Initially, we used several nanocharacterization techniques to analyze the size of vesicles and visualize these particles in CFE extracts prepared using homogenization and 30,000 *x g* clarification (i.e., S30 extracts) as described previously[34]. Dynamic light scattering (DLS) analysis of crude extract revealed two major peaks: one narrower peak with an intensity maximum at ~20 nm, and a broader peak at ~100-200 nm (**Fig. 2A**). The 20 nm peak likely represents small cell-derived particles. *E. coli* ribosomes, which are present at ~1 μM in typical CFE reactions, and enabled the production of sfGFP in our CFE reactions (**Supplementary Fig. 1**), are ~20 nm in size and likely contribute considerably to the signal measured[73,74]. We hypothesized that particles measured in the ~100-200 nm peak were vesicles. An illustration of particles detected in extract is shown in **Fig. 2C**. To directly analyze membrane vesicles without ribosomes and other cellular particles, we identified and purified membranous particles via size

exclusion chromatography (SEC)[75–77] (**Supplementary Fig. 2A**). DLS analysis of purified membrane vesicles revealed an intensity particle size distribution that directly overlapped with the proposed vesicle peak from our DLS traces of crude extracts (**Fig. 2A**). Nanoparticle Tracking Analysis (NTA), an orthogonal method for sizing and quantitating nanoparticles in solution, revealed an average purified vesicle diameter of 118.5±0.7 nm, corroborating the approximate size range of vesicles measured with DLS (**Fig. 2B**). The zeta potential of purified vesicles was -14.5±-1.0 mV, indicating a negative particle surface charge consistent with phospholipid vesicles (**Supplementary Fig. 2B**). Cryo-electron microscopy (cryo-EM) of extracts showed small (≤20 nm) particles and other larger, circular particles consistent with vesicle morphology (**Fig. 2D**). Cryo-EM micrographs of extracts revealed vesicles between ~40 nm and ~150 nm in size, and we observed intact vesicle morphologies both pre- and post- SEC purification (**Fig. 2D-2E**). Uncropped and annotated Cryo-EM micrographs are shown in (**Supplementary Fig. 3)**. Comparisons between measurements reveal that DLS, a bulk, in-solution measurement, over-estimates vesicle diameter. DLS, however, is a useful tool for quickly characterizing crude extract particle profiles because it can detect particles <50 nm (including ribosomes) that are smaller than vesicles and are below the size limit of detection of NTA. Together, these results show particle profiles of crude extracts and reveal that vesicles are polydisperse, are on the order of tens to hundreds of nm across, and are relatively low in concentration compared with ribosomes and other small complexes.

***Extract processing impacts vesicle size distributions and concentrations***

To understand how to control membrane vesicles in extracts, we next sought to study how protocols to process extracts impacted vesicle properties. Specifically, we studied cell lysis and extract centrifugation because cell membranes are ruptured during lysis, and centrifugation dictates particle separation. We lysed cells using standard sonication (constant input energy per volume of cell suspension) or homogenization protocols (~20,000 psig)[11,34], then subjected

lysates to a traditional 30,000 x *g* centrifugation protocol (termed 'S30' extracts), or a lower g-force protocol where the maximum centrifugation speed was 12,000 x *g* (termed 'S12' extracts) (**Fig. 3A**)[11,12]. These combinations of lysis and centrifugation protocols were selected because they have previously been used to obtain high-yielding *E. coli* CFE extracts.[78] Indeed, all the conditions tested yielded extracts that were active for protein synthesis in standard CFE reaction conditions (**Supplementary Fig. 4A**). The combination of a standard homogenization and S30 prep represents our base case because extracts used in our previously described one-pot cell-free glycoprotein synthesis platform were prepared with these conditions, as well as the extracts used in **Fig. 2**. Before this work, S12 extracts had not previously been used for making glyco-competent CFE extracts.

Of the conditions tested, the centrifugation protocol had the most impact on vesicle concentrations. We observed significantly higher numbers of vesicles in S12 extracts for both lysis methods, with the reduced centrifugation speed likely being the reason for increased particle concentrations. Specifically, we observed 1.2- and 2.0- fold enrichments of vesicles in sonicated and homogenized S12 extracts, respectively (**Fig. 3B**). Homogenized S12 extracts contained the highest concentration of vesicles with $6.5 \pm 0.3 \times 10^{12}$ particles/mL (as compared to $3.4 \pm 0.1 \times 10^{12}$ particles in the base case), making it the most promising condition for enriching vesicles.

While centrifugation impacted vesicle concentration, lysis method impacted vesicle size. Sonicated extracts contained smaller vesicles with narrower size distributions than homogenized extracts, regardless of centrifugation protocol. Our observations that lysis method impacts vesicle size is consistent with studies showing that varying experimental parameters to disperse phospholipids (or amphiphiles in general) impacts vesicle sizes[79]. Particle size distributions of sonicated extracts reached single maxima at ~110 nm, with average particle diameters of ~130nm; homogenized extracts had higher average particle diameters of ~160 nm, displaying distinct peaks at ~120 nm, and considerable shoulder peaks at ~150 nm (**Fig. 3C-3D,**

**Supplementary Fig. 5A**). The particle size distributions observed in homogenized extracts may indicate the presence of multiple, discrete, vesicle populations (**Fig 3C-3D)**. DLS measurements confirmed the observation that sonicated extracts contained relatively smaller, less polydisperse vesicles than homogenized extracts (**Supplementary Fig. 5B-5C**). Notably, direct vesicle analysis in extracts enabled us to gauge the impacts of extract processing in ways that have not been previously accessible and provides benchmarks for intact vesicle concentrations in extracts.

***Heterologous membrane-bound cargo can be controllably enriched via membrane vesicles***

With a better understanding of the characteristics and concentrations of native vesicles, we sought to enrich extracts with vesicles containing heterologous cargo derived from the periplasmic membrane of *E. coli.* Since S12 extracts contain higher concentrations of vesicles than S30 extracts, we hypothesized that S12 extracts would also contain higher concentrations of associated heterologous cargo. The highest dynamic range of vesicle concentration between S12 and S30 preps was observed with homogenization, so we proceeded with homogenization for enrichment experiments (**Fig. 3B**). We overexpressed six membrane-bound proteins of various sizes, transmembrane topologies, biological functions, and taxonomical origins to test for enrichment (**Supplementary Table 1**). The proteins selected for enrichment encompass classes of proteins that could enable new functionalities in CFE, including glycosylation enzymes (PglB, PglO, STT3) and signal transduction/sensing proteins (NarX, PR, CB1). We expressed each membrane protein *in vivo* with a C-terminal FLAG tag, prepared S30 and S12 extracts, then analyzed concentrations of the overexpressed membrane protein using quantitative Western blotting. We observed approximately 2-fold membrane protein enrichment in S12 over S30 (S12/S30) extracts for all proteins other than PR, for which we observed 4-fold enrichment (**Fig. 4A-4B**). As a control, when sfGFP with no transmembrane helices was

expressed *in vivo*, we did not observe significant S12/S30 enrichment (**Fig. 4C**). Full blots for **Fig. 4A-4C** are shown in **Supplementary Fig. 6**. Notably, enrichment values obtained via blotting correspond closely with the 2-fold vesicle enrichment observed via NTA in homogenized S12 and S30 extracts with no overexpression (**Fig. 3B**). All extracts with pre-enriched membrane proteins displayed protein synthesis activity (**Supplementary Fig. 4B**).

With an eye towards bacterial glycoengineering applications, we next confirmed that PglB and PglO, key enzymes for glycosylation, were associated with membrane vesicles, as opposed to free in solution (**Fig. 4D**). Extracts with pre-enriched PglB or PglO were probed with a green fluorescent α-FLAG antibody, then analyzed via SEC. Fluorescence chromatograms are shown in **Fig. 4D**, with the characteristic vesicle elution fraction highlighted in gray (**Supplementary Fig. 2A**). The characteristic vesicle elution peak corresponded with green fluorescence for extracts containing PglB or PglO and no corresponding peak was observed in an extract with no overexpressed membrane protein (**Fig. 4D**). Our results show that heterologous cargo that is embedded in the periplasmic (inner) membrane of *E. coli* cells can be pre-enriched in extract and tuned via vesicles.

### *Increasing vesicle concentrations improves cell-free glycoprotein synthesis (CFGpS) for N- and O-linked glycosylation systems*

We next set out to exploit our ability to enrich vesicles harboring heterologous cargo in an application. We focused on protein glycosylation, because glycosylation plays critical roles in cellular function, human health, and biotechnology. As a model, we sought to increase glycoprotein yields in a previously reported CFGpS platform by charging reactions with S12 extracts containing higher concentrations of membrane-bound glycosylation machinery[34]. We prepared S30 and S12 extracts from strains overexpressing the model *N*-linked glycosylation pathway from *Campylobacter jejuni*, which consists of the membrane-bound oligosaccharyltransferase (OST) PglB that catalyzes glycosylation, and a lipid-linked

oligosaccharide (LLO) donor of the form: GalNAc-α1,4-GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac (where Bac is 2,4-diacetamido-2,4,6-trideoxyglucopyranose) from an undecaprenylpyrophosphate-linked donor[80]. NTA and Western blot analysis of CFGpS extracts revealed 2.5-fold S12/S30 enrichment of vesicles and a corresponding 2-fold S12/S30 enrichment of PglB (**Supplementary Fig. 7**). Fluorescence staining and SEC analysis confirmed the presence and association of LLO and PglB with the vesicles (**Supplementary Fig. 8A**).

To assess the impact of enriched vesicles on cell-free glycoprotein synthesis, we carried out reactions in two phases (**Fig. 5A, Inset**)[24]. First, cell-free protein synthesis (CFPS) of the acceptor protein was run for a defined time, termed 'CFPS time'. At the CFPS time, reactions were spiked with $MnCl_2$, quenching CFPS and initiating glycosylation by providing the OST with its $Mn^{2+}$ cofactor. CFGpS reactions charged with S30 or S12 extracts were run for CFPS times of 2, 10, 20, 30, and 60 minutes using a His-tagged $sfGFP_{DQNAT}$ acceptor protein, where DQNAT is a permissible PglB sequon (with N being the glycosylated residue). Coding sequences of all acceptor proteins used are presented in **Supplementary Table 2**. Endpoint glycoprotein yields were quantified using total acceptor protein yield and % glycosylation determined by Western blotting (**Fig. 5A, Supplementary Fig. 9A-9D)**. At longer CFPS times we observed that S12 extracts produced significantly more glycoprotein than S30 extracts. Because total acceptor protein concentrations for S30 and S12 reactions were similar for each CFPS time (**Supplementary Fig. 9E)**, increased glycoprotein yield in S12 extracts is due to higher glycosylation activity and not higher CFE yields**.** Specifically, at 20, 30, and 60-minute CFPS times, we observed 67%, 85%, and 91% increases in glycoprotein yield in the S12 reactions, respectively. At the 60-minute CFPS time, S12 reactions yielded 117.2±9.9 μg/mL of glycoprotein in batch **(Fig. 5A)**. To our knowledge, this is the first time that batch glycoprotein titers on the order of hundreds of μg/mL have been synthesized in a crude-extract-based CFGpS system without extra vesicle supplementation to the reactions[2]. This advance was

enabled by using S12 extracts instead of S30 extracts, and relying on enriched cell-derived membranes. S12 reactions also had significantly higher terminal % glycosylation, or percent of CFPS-derived acceptor protein that is glycosylated at the end of a 16-hour glycosylation reaction. This was true for all CFPS times tested (**Supplementary Fig. 9F**). For example, for reactions with 20-minute CFPS times, we observed an increase from 51% glycosylation for S30 reactions to 82% glycosylation for S12 reactions (**Fig. 5B, Supplementary Fig. 9F**). α-His (showing glycosylated and aglycosylated acceptor protein) and α-glycan (against the *C. jejuni* glycan) Western blots of representative reactions are shown in **Fig. 5C.** Taken together, these results indicate that the higher concentrations of membrane associated glycosylation components in S12 extracts has a measurable effect on CFGpS, improving glycoprotein yields and endpoint % glycosylation.

With a long-term interest in synthesizing diverse glycoproteins in cell-free systems, we next ported an *O*-linked glycosylation system, known to have broad glycan specificity, into the CFGpS platform[72,81,82]. We selected the *O*-OST PglO from *Neisseria gonorrhoeae* which accepts the *C. jejuni* heptasaccharide LLO as a donor but differs from PglB in acceptor sequence preferences[83]. For PglO, we used an sfGFP-fusion acceptor protein containing a recently determined 8 amino acid (WPAAASAP, with S being the glycosylated residue) minimum optimal *O*-linked recognition site (termed 'MOOR').[83] We confirmed residue-specific *O*-linked glycosylation and enrichment of PglO and LLO in vesicles (**Supplementary Fig. 10, Supplementary Fig. 8B**)**.** As additional proof of site-specific glycosylation, we performed LC-MS/MS analysis of the glycoproteins obtained via CFGpS with PglO and PglB and observed the presence of the 1406 Da *C. jejuni* heptasaccharide on the expected tryptic peptides (**Supplementary Figure 11A-11B**)[80]. As in PglB-mediated CFGpS, we observed increased endpoint glycoprotein yield and % glycosylation in reactions charged with S12 extracts. Specifically, reactions with CFPS times of 20 minutes resulted in a 69% increase in glycoprotein yield and an increase from 27% to 40% glycosylation in reactions with S12 extracts compared to

1 those containing S30 extracts (**Fig. 5D**, **Supplementary Fig. 12**)**.** Corresponding blots are
2 shown in **Fig. 5E** and **Supplementary Fig. 12A-12B.** Taken together, these results indicate that
3 improvements to glycosylation in S12 extracts translate from the *N*-linked glycosylation system
4 to the *O*-linked glycosylation system.

5 To determine whether enhanced glycoprotein production in S12 extract-based CFGpS
6 reactions was transferrable to non-model acceptor proteins, we tested three additional proteins.
7 This included the *Campylobacter jejuni* AcrA, a native bacterial glycoprotein with two internal
8 glycosylation sites[50,84], as well as two possible carrier proteins for conjugate vaccines. The
9 possible carrier proteins were: *H. influenzae* protein D (PD), which is a licensed carrier protein,
10 and *E. coli* maltose binding protein (MBP), which is not yet a licensed carrier but has shown
11 promising results in clinical studies[85,86]. PD and MBP were each fused to a single C-terminal
12 DQNAT sequon to enable glycosylation[24]. CFGpS was run as previously described using a 20
13 min CFPS time, with all conditions held constant (other than the DNA template and $^{14}$C-leucine to
14 enable quantification, supplemented to the reactions). We observed 80%, 147%, and 167%
15 increases in glycoprotein titers for AcrA, MBP, and PD, respectively, when comparing S12- to
16 S30-based reactions. Expression improvements were determined by $^{14}$C-leucine incorporation
17 (**Fig. 5F**, **Supplementary Fig. 13A-13E**). Our results highlight that the improvements in
18 glycoprotein yield observed in extracts with higher concentrations of vesicles hold for diverse
19 proteins without the need for re-optimization.
20
21 **Discussion**
22 In this work, we set out to benchmark, understand, and quality-control protein-enriched
23 vesicles in bacterial CFE extracts for expanding and enhancing functionality. We showed that
24 upstream extract processing can be used to tune concentrations of vesicles and associated cargo
25 from the periplasm. Then, we applied this knowledge to improve cell-free glycoprotein synthesis,
26 with a specific application focus of glycoprotein synthesis. Our results have several key features.

First, the light scattering tools used here allowed us to quickly quantify intact vesicle numbers and sizes in CFE extracts. This is important because this knowledge informed design rules for enhancing vesicle concentrations and functionality from their associated protein cargo in cell-free systems. Notably, the effective vesicle surface area calculated from NTA measurements (~0.3 $m^2$ membrane/mL extract), is consistent with values calculated from phospholipid concentrations in similar extracts[56].

Second, our results offer insights into field-wide observations and limitations of *E. coli*-based CFE systems. For example, it is well-documented that lysis protocols can profoundly impact CFE productivity[87]. Our findings show that lysis methods impact size distributions of vesicles generated during this step which affect the membrane environment of the machinery necessary for oxidative phosphorylation and ATP regeneration. Since vesicles are key for activating cost-effective energy metabolism from oxidative phosphorylation in CFE, routine vesicle characterization could become a vital quality-control check, leading to improved reproducibility in and between labs[88]. Our results also offer insight into why, despite the presence of vesicles in the *E. coli* CFE system, CFE-derived membrane proteins cannot be synthesized via insertion into native vesicles without additional vesicle supplementation[48,50,54]. With ~6 nM of intact vesicles in CFE reactions (where intact vesicle concentration was calculated from NTA measurements), the concentration of vesicles is orders of magnitude lower than typical protein titers produced in our CFE extracts (~30 μM of reporter protein or higher).

Third, our work opens the door to engineering cell-free systems that rely on enriched membrane-bound components. We show that membrane-bound proteins and lipid-linked oligosaccharides expressed *in vivo* in the periplasm can be enriched in vesicles, indicating that a population of vesicles is derived from the inner periplasmic membrane[89,90]. Importantly, our workflow easily interfaces with numerous methods that could be used to alter vesicles and their membrane-bound cargo. For example, using other centrifugation speeds besides 12,000 x *g* could result in changes to vesicle concentration. In addition, additives could be supplemented to

cell-free systems to tune biophysical features of membrane properties (e.g., composition, size, fluidity, curvature). Furthermore, unlike the previously used S30 extract procedure, the optimized S12 extract strategy developed here does not require a high-speed centrifuge and is less time-intensive. This simplifies the CFGpS platform, enabling the process from inoculation of cell culture to testing CFGpS reactions to be completed in a single workday. And, while we focus entirely on *E. coli*-based systems here, the reported characterization methods could, in principle, be extended to further optimize insect and CHO-based CFE systems that rely on ER-derived microsomes to perform glycosylation, embed nascent membrane proteins, and perform other membrane-dependent functions.

Towards applications in biomanufacturing, a key feature of the *E. coli*-based CFGpS system is expressing synthetic glycosylation pathways encoding diverse O-antigens from pathogenic bacteria. This feature points toward immediate utility of our CFGpS system in the on-demand bioproduction of conjugate vaccines[24]. Here, we show that S12 extracts enable higher glycoprotein titers of two glycoconjugate vaccine carrier proteins modified with a model *C. jejuni* LLO, indicating that vaccine production may be simpler and more efficient using the optimized methods reported here. Additionally, we have recently shown that our optimized S12 conditions can be used to recapitulate efficient, humanized *O*-linked glycosylation in glycoengineered E. coli extracts[72]. While applications in *O*-linked glycosylation and conjugate vaccines are imminent, the recapitulation of efficient eukaryotic-type *N*-linked glycosylation (i.e., glycoproteins with a $Man_3GlcNAc_2$ core glycan) for therapeutics production still remains on the horizon in *E. coli*-based systems.

Future studies to elucidate translocation and co-translational glycosylation in vesicles will be important. These studies could be especially useful for producing complex, native glycoproteins for which protein glycosylation and folding are co-translational. While it has been shown that glycosylation with PglB can proceed on pre-folded proteins *in vitro* (using purified, reconstituted components and without the need for translocation or intact membranes[91]),

obtaining a more robust understanding of the topology of glycosylation in the membrane vesicles extracts is an important future effort for therapeutics production.

Looking forward, we anticipate that our work will accelerate efforts to manufacture proteins that require membrane-dependent modifications, such as glycoproteins. For example, the approach described enables *N*-linked glycoprotein synthesis yields of >100 µg/mL, which increases accessibility for on-demand vaccine production in resource-limited settings. Taken together, our results pave the way for efficient, accessible CFE systems that require membrane-bound activities for expanding system functionality and enabling a variety of synthetic biology applications.

## Methods

### *Extract preparation*

The chassis strain used for all extracts was CLM24[34]. Source strains were grown in 1 L of 2×YTPG media at 37 °C with agitation. Cells were grown to OD 3, then harvested by centrifugation (5000 × *g*, 4 °C, 15 min). For overexpression of proteins *in vivo*, CLM24 source strains were grown at 37 °C in 2xYTPG with the appropriate antibiotic(s), listed in **Supplementary Table 3.** Cells were induced with 0.02% (w/v%) L-arabinose at OD 0.6-0.8, shifted to 30 °C, and harvested at OD 3. All subsequent steps were carried out at 4 °C and on ice unless otherwise stated. Pelleted cells were washed 3 times in S30 buffer (10 mM Tris acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate). After the last wash, cells were pelleted at 7000 × *g* for 10 min, flash-frozen and stored at -80 °C. After growth and harvest, cells were thawed and resuspended to homogeneity in 1 mL of S30 buffer per gram of wet cell mass. For homogenization, cells were disrupted using an Avestin EmulsiFlex-B15 high-pressure homogenizer at 20,000-25,000 psig with a single pass (Avestin, Inc. Ottawa, ON, Canada). For sonication, input energy was calculated using an empirical correlation as

described previously [11]. Cells were sonicated on ice using a Q125 Sonicator (Qsonica, Newtown, CT) with a 3.175 mm diameter probe at a frequency of 20 kHz and 50% of amplitude. Energy was delivered to cells in pulses of 45 s followed by 59 s off until the target energy was delivered. Cells were lysed and clarified in triplicate. For S30 prep, lysed cells were centrifuged twice at 30,000 × *g* for 30 min; supernatant was transferred to a fresh tube for each spin. Supernatants were incubated with 250 rpm shaking at 37 °C for 60 min for runoff reactions. Following runoff, lysates were centrifuged at 15,000 × *g* for 15 min. Supernatants were collected, aliquoted, flash-frozen, and stored at −80 °C for further use. For S12 prep, lysed cells were centrifuged once at 12,000 *x g* for 10 min; supernatants were collected and subjected to runoff reactions as described above. Following runoff, lysates were centrifuged at 10,000 × *g* for 10 min at 4 °C. Supernatants were collected, aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

***Dynamic light scattering (DLS) and nanoparticle tracking analysis (NTA) measurements***

DLS measurements were performed on a Zetasizer Nano ZS (Malvern Instruments Ltd., UK) with a measurement angle of 173$^{o}$ in disposable cuvettes (Malvern Instruments Ltd., UK ZEN0040). All measurements were collected in triplicate for 13 scans per measurement. Refractive index and viscosity were obtained from the instrument's parameter library. The instrument's 'General Purpose' setting was used to calculate intensity and number particle size distributions. For DLS of crude extracts, extracts were diluted 1:10 with 0.1 µm filtered PBS before analysis. For purified vesicle samples, elutions were analyzed directly without dilution.

NTA measurements were performed on a Nanosight NS300 using a 642 nm red laser (Malvern Instruments Ltd., UK). Samples were diluted to manufacturer-recommended particle concentrations in sterile PBS until a linear trend between dilution factor and concentration measured was found. Samples were flowed into the cell, and the instrument was focused

according to manufacturer recommendations. Measurements were collected at room temperature, using a 1 mL syringe and a syringe pump infusion rate of 30 (arbitrary units). Data for each sample was collected in 5 separate 1 min videos, under continuous flow conditions. Mean particle diameters and particle concentrations were obtained from aggregate Nanosight experiment reports of each run, then averaged across triplicates and corrected for dilution factor.

***Transmission electron microscopy***

For cryo-TEM measurement, 200 mesh Cu grids with a lacey carbon membrane (EMS Cat. # LC200-CU) were placed in a Pelco easiGlow glow discharger (Ted Pella Inc., Redding, CA, USA) and an atmosphere plasma was introduced on the surface of the grids for 30 seconds with a current of 15 mA at a pressure of 0.24 mbar. This treatment creates a negative charge on the carbon membrane, allowing for aqueous liquid samples to spread evenly over of the grid. 4 µL of sample was pipetted onto the grid and blotted for 5 seconds with a blot offset of +0.5 mm, followed by immediate plunging into liquid ethane within a FEI Vitrobot Mark III plunge freezing instrument (Thermo Fisher Scientific, Waltham, MA, USA). Grids were then transferred to liquid nitrogen for storage. The plunge-frozen grids were kept vitreous at –172 °C in a Gatan Cryo Transfer Holder model 626.6 (Gatan Inc., Pleasanton, CA, USA) while viewing in a JEOL JEM1230 LaB6 emission TEM (JEOL USA, Inc., Peabody, MA,) at 120 keV. Image data was collected by a Gatan Orius SC1000 CCD camera Model 831 (Gatan Inc., Pleasanton, CA, USA). Image analysis was done using Image J.

***Western blotting and densitometry analyses***

SDS-PAGE was run using NuPAGE 4-12% Bis-Tris protein gels with MOPS-SDS buffer (Thermo Fisher Scientific, USA). After electrophoresis, proteins were transferred from gels to Immobilon-P polyvinylidene difluoride 0.45 µm membranes (Millipore, USA) according to

manufacturer's protocol. Membranes were blocked in either Odyssey or Intercept blocking buffer (LI-COR, USA). α-FLAG blots of membrane proteins were probed using α-FLAG antibody (Abcam 2493) as the primary. α-His blots were probed with 6xHis-antibody (Abcam, ab1187) as the primary. For α-glycan blots, hR6 serum from rabbit that binds to the native *C. jejuni* glycan was used as the primary probe [49]. A fluorescent goat α-Rabbit IgG IRDye 680RD (LI-COR, USA) was used as the secondary for all blots. Blots were imaged using a LI-COR Odyssey Fc (LI-COR Biosciences, USA). Densitometry was preformed using Image Studio Lite software to measure band intensity. Fluorescence background was subtracted from blots before determining band intensities. For determining membrane protein enrichment (S12/S30), band intensities of membrane proteins for three independent S12 extract replicates and three independent S30 replicates were measured for each protein. The rounded averages of triplicate ratios (S12/S30) and associated error are reported as enrichment in **Fig. 3.** For determining glycoprotein yields from CFGpS reactions, band intensities for glycosylated and aglycosylated bands were obtained from independent, triplicate reactions. The fraction of glycosylated protein for each replicate was calculated via band intensities. To obtain glycoprotein yields, the fraction glycosylated was multiplied by total protein yield for each replicate as calculated from sfGFP fluorescence converted to protein concentration (described below).

***Lipid dye staining and fluorescence immunostaining of vesicles***

All reagents used for immunostaining and SEC were sterile filtered with a 0.1 μm filter (Millex-VV Syringe Filter, Merck Millipore Ltd. or Rapid-Flow Filter, Nalgene). To determine vesicle elution fractions, extract was probed with FM 4-64 lipid dye (Life Technologies), a lipophilic styrene dye that has low fluorescence in aqueous solution and becomes brightly fluorescent upon incorporation into membranes. FM-464 dye preferentially stains the inner membrane of *E. coli,* but has been used to dye the outer membrane as well[92,93]. FM 4-64 lipid

dye was prepared in stock solutions at 10 mg/mL in 100% DMSO, then diluted 1,000-fold in nuclease free water before use. 80 µL of extract, 10 µL 10x PBS, and 10 µL of FM 4-64 were mixed to a final concentration of 1 ng dye/µL. Samples were incubated with dye in the dark for 10 mins at 37 °C prior to SEC. To verify the presence of glycosylation components in vesicles, we probed for the LLO with a red fluorescent soybean agglutinin (SBA) lectin, a protein complex which specifically binds to the *C. jejuni* LLO[80], and for PglB with an orthogonal green fluorescent α-FLAG antibody as described above. For α-FLAG immunostaining and SBA staining, 90 µL extract and 10 µL of 10xPBS were mixed with 2 µL of α-FLAG-DyLight 488 (Invitrogen, MA191878D488) and 4 µL of SBA-AlexaFluor™ 594 (Invitrogen, 32462). Antibody and SBA were incubated with extract in the dark with agitation overnight at 4 °C prior to SEC.

***Size exclusion chromatography (SEC) of vesicles***

100 µL of extract mixture (stained with lipid dye or antibody) was flowed over a size exclusion chromatography column with PBS. Elution fractions were collected into a clear polystyrene 96-well plate (Costar 3370, Corning Inc., USA) at a rate of 0.4 min/well using a Gilson FC 204 Fraction Collector (Gilson, Inc., USA). Poly-Prep chromatography columns (Bio-Rad, USA) were packed with 8 mL of Sepharose 4B resin 45-165 µm bead diameter, (Sigma Aldrich, USA) and washed with sterile PBS 3 times before use. This resin was chosen based on the ~100 nm particle size observed in scattering measurements. Elution fluorescence was measured using a Synergy H1 microplate reader (Biotek, USA). Excitation and emission wavelengths for SBA-AlexaFluor™ 594 were 590 and 617 nm, respectively. Excitation and emission wavelengths for α-FLAG-DyLight 488 were 493 and 528 nm, respectively. Vesicles stained with FM 4-64 lipid dye were used to determine the characteristic vesicle elution fraction. Reference samples probed with FM 4-64 were used to determine the characteristic vesicle

elution fraction in each experiment. For plots, each curve was background subtracted and normalized to the highest RFUs measured for each respective fluorescent elution profile.

***CFE reactions***

Protein synthesis was carried out with a modified PANOx-SP system in triplicate reactions, with each reaction containing a uniquely-prepared extract[87]. Specifically, 1.5 mL microcentrifuge tubes (Axygen, MCT-150-C) were charged with 15 µL reactions containing 200 ng pJL1-sfGFP plasmid (**Supplementary Table 1),** 30% (v/v%) extract and the following: 6 mM magnesium glutamate (Sigma, 49605), 10 mM ammonium glutamate (MP, 02180595), 130 mM potassium glutamate (Sigma, G1501), 1.2 mM adenosine triphosphate (Sigma A2383), 0.85 mM guanosine triphosphate (Sigma, G8877), 0.85 mM uridine triphosphate (Sigma U6625), 0.85 mM cytidine triphosphate (Sigma, C1506), 0.034 mg/mL folinic acid, 0.171 mg/mL *E. coli* tRNA (Roche 10108294001), 2 mM each of 20 amino acids, 30 mM phosphoenolpyruvate (PEP, Roche 10108294001), 0.4 mM nicotinamide adenine dinucleotide (Sigma N8535-15VL), 0.27 mM coenzyme-A (Sigma C3144), 4 mM oxalic acid (Sigma, PO963), 1 mM putrescine (Sigma, P5780), 1.5 mM spermidine (Sigma, S2626), and 57 mM HEPES (Sigma, H3375). To gauge extract CFE productivity, reactions were carried out for 20 hours at 30 °C.

***Quantification of CFE and CFGpS protein yields***

As described previously, the concentration of cell-free-derived sfGFP was determined by measuring in-extract fluorescence and then converting to protein concentration using a standard curve relating sfGFP fluorescence to protein concentration as determined by a [$^{14}$C]-leucine incorporation assay[34]. Briefly, 2 µL of cell-free reaction product was diluted into 48 µL of Ambion nanopure water (Invitrogen, USA). The solution was then placed in a Costar 96-well black assay plate (Corning, USA). Fluorescence was measured using a Synergy H1 microplate reader

(Biotek, USA). Excitation and emission wavelengths for sfGFP fluorescence were 485 and 528 nm, respectively. This RFU value was then used to calculate the protein concentration.

Yields of all acceptor proteins (other than sfGFP) were assessed directly via the addition of 10 μM [$^{14}$C]-leucine (PerkinElmer) to the CFGpS reaction to yield trichloroacetic acid-precipitable radioactivity that was measured using scintillation counting. Soluble fractions were isolated after centrifugation at ≥ 12,000 × *g* for 15 min at 4 °C. Briefly, 6 μL of the soluble fraction of CFGpS reactions run with 20 min CFPS times were mixed with 6 μL 0.5 M KOH and incubated for 20 min at 37°C. 5 μL of treated sample was then soaked into 2 separate filtermats (PerkinElmer Printer Filtermat A 1450-421) and dried under a heat lamp. One filtermat was washed three times using 5% trichloroacetic acid (TCA) with 15 min incubations at 4°C, and then once with ethanol with a 10 min incubation at room temperature. Following melting of scintillation wax (PerkinElmer MeltiLex A 1450-441) on top of both TCA-precipitated and non-TCA precipitated filtermats, incorporated radioactivity was measured by a Microbeta2 (PerkinElmer) scintillation counter. Low levels of background radioactivity in S12 and S30 extracts were measured in CFGpS reactions containing no plasmid DNA template and subtracted before calculation of protein yields. The fraction of incorporated leucine (washed/unwashed) was used to determine the amount of protein produced in each reaction.

***Autoradiograms of CFGpS reaction products***

For sfGFP-based glycosylation experiments, Western blotting of the acceptor proteins followed by densitometry analysis was used to quantitate the fraction of acceptor protein glycosylated (see above for detailed description). For other acceptor proteins, autoradiograms were used to quantitate glycoprotein from CFGpS reaction products using densitometry. Autoradiograms were run by first running SDS-PAGE gels of the soluble fractions of CFGpS reactions (from the

same reactions used to calculate yields) using NuPAGE 4-12% Bis-Tris protein gels with MOPS-SDS buffer (Thermo Fisher Scientific, USA). The gels were then dried overnight between cellophane films and then exposed for 48-72 hours to a Storage Phosphor Screen (GE Healthcare). The Phosphor Screen was imaged using a Typhoon FLA7000 imager (GE Healthcare). Autoradiogram gel images were acquired using Typhoon FLA 7000 Control Software Version 1.2 Build 1.2.1.93. Autoradiogram analysis was performed using ImageJ (Version 2.1.0/1.53c, Build 5f23140693) gel analyzer to determine ratios of glycosylated and aglycosylated full-length acceptor protein. Glycoprotein yields were determined by multiplying fraction glycosylated as determined by ImageJ analysis, by the yields determined from scintillation counting for each replicate.

***Cell-free glycoprotein synthesis (CFGpS) reactions***

For crude extract-based expression of glycoproteins, a two-phase scheme was implemented as previously described[34]. In this work, protein synthesis was carried out as described above at 15 µL in PCR strip tubes (Thermo Scientific AB-2000) with 50 ng template DNA. Reactions were supplemented with the plasmids encoding permissible or non-permissible sequons on sfGFP acceptor proteins. pJL1-sfGFP-DQNAT-His (permissible) and pJL1-sfGFP-AQNAT-His (non-permissible) were used for PglB-mediated glycosylation; pJL1-sfGFP-MOOR-His (permissible) and pJL1-sfGFP-$MOOR_{mut}$-His (non-permissible) were used for PglO-mediated glycosylation (**Supplementary Table 2**). Reactions were set up in triplicate on ice, with each reaction containing a uniquely-prepared extract. CFPS time was measured as the time at which reactions were moved to 30 °C to the time when reactions were spiked with $MnCl_2$. In the second phase, protein glycosylation was initiated by the addition of $MnCl_2$ at a final concentration of 25 mM. In addition to $MnCl_2$ (Sigma 63535), either 0.1% (w/v%) DDM (Anatrace, D310S) or 100 mM sucrose was supplemented to PglB or PglO reactions,

respectively. Glycosylation proceeded at 30 °C for 16 hrs. After glycosylation, GFP fluorescence was used to quantitate the total amount of acceptor protein synthesized, and Western blots were used to calculate the fraction of glycosylated and aglycosylated proteins.

***Estimation of vesicle membrane area***

The equation below was used to calculate vesicle surface area ($m^2$/mL), where $R_{ave}$ is average vesicle radius (m), C is concentration of particles measured by NTA (particles/mL).

$$vesicle\ surface\ area = 4 * \pi * (R_{ave})^2 * C$$

***Liquid Chromatography Mass Spectrometry (LC-MS/MS)***

Acceptor proteins were purified using a His purification protocol prior to LC-MS. CFGpS reactions producing glycosylated sfGFP-DQNAT and sfGFP-MOOR were scaled up to a total volume of 1.2 mL each and run in 50 mL conical tubes (Falcon, Corning) with 20-minute CFPS times. Following 16-hour glycosylation reactions, CFGpS reactions were transferred to 1.5 mL microtubes (Axygen, Corning) and centrifuged at 16,000 x g for 3 minutes. Soluble fractions were split in half and loaded onto 2 equilibrated Ni-NTA Spin Columns (Qiagen 31014) per CFGpS reaction following column equilibration with equilibration buffer (50 mM $NaH_2PO_4$, 300 mM NaCl and 10 mM imidazole). CFGpS reactions were incubated on columns for 5 minutes at room temperature followed by centrifugation at 250 x g for 12 minutes. Columns were then washed 3 times with 600 μL low imidazole buffer (50 mM $NaH_2PO_4$ and 300 mM NaCl and 20 mM imidazole) and centrifuged at 900 x g for 2 minutes before elution in 100 μL of high-imidazole buffer (50 mM $NaH_2PO_4$ and 300 mM NaCl and 500 mM imidazole). Four elution fractions were collected, and the most concentrated fraction collected from each column was dialyzed against 50 mM Ammonium Bicarbonate. Dialysis buffer was changed after 2 hours and then allowed to proceed overnight.

Glycopeptides for LC-MS/MS analysis were prepared by reducing His-tag purified, dialyzed glycosylation targets by incubation with 5 mM DTT at 60 °C for 1 hour and then digesting with 0.0044 µg/µl MS Grade Trypsin (Thermo Fisher Scientific) at 37 °C overnight. LC-MS/MS was performed by injection of 20 µl (or about 35 pmol for sfGFP-DQNAT and 25 pmol for sfGFP-MOOR) of digested glycopeptides into a Bruker Elute UPLC equipped with an ACQUITY UPLC Peptide BEH C18 Column, 300 Å, 1.7 µm, 2.1 mm × 100 mm (186003686 Waters Corp.) with a 10 mm guard column of identical packing (186004629 Waters Corp.) coupled to an Impact-II UHR TOF Mass Spectrometer. As described previously, liquid chromatography was performed using 100% $H_2O$ and 0.1% formic acid as Solvent A and 100% acetonitrile and 0.1% formic acid as Solvent B at a flow rate of 0.5 mL/min and a 40 °C column temperature. An initial condition of 0% B was held for 1 min before elution of the peptides of interest during a 4 min gradient to 50% B. The column was washed and equilibrated by a 0.1 min gradient to 100% B, a 2 min wash at 100% B, a 0.1 min gradient to 0% B, and then a 1.8 min hold at 0% B, giving a total 9 min run time[17]. Pseudo multiple reaction monitoring (MRM) MS/MS fragmentation was targeted to theoretical glycopeptide masses corresponding to detected peptide MS peaks. Glycopeptides were fragmented with a collisional energy of 30 eV and an isolation window that included the entire glycopeptide isotopic envelope. For LC-MS/MS of glycopeptides, a scan range of 100–3000 *m/z* with a spectral rate of 8 Hz was used. External calibration was performed prior to data collection.

LC-MS(/MS) data was collected using Bruker Compass Hystar v5.0 and analyzed using Bruker Compass Data Analysis v4.4 (Bruker Daltonics, Inc.). Representative LC-MS/MS spectra from MRM fragmentation were selected and annotated manually. Observed glycan and glycopeptide *m/z* values are annotated in figures. LC-MS/MS data was exported from Bruker Compass Data Analysis and plotted in Microsoft Excel.

**Acknowledgments**

We acknowledge Jessica Stark and Weston Kightlinger for helpful discussions about CFGpS, Aravind Natarajan and Dominic Mills for helpful discussions about *O*-OSTs and their substrate specificities, Ashty Karim for helpful discussions on data visualization, Han Teng Wong and Charlotte Abrahamson for fruitful discussions regarding immunostaining and Maddie DeWinter for helpful discussions on LC-MS/MS methods. We thank Markus Aebi for providing the hR6 serum and Jeff Tabor for providing the DNA encoding NarX. This work made use of the BioCryo facility of Northwestern University's NUANCE Center, which has received support from the SHyNE Resource (NSF ECCS-1542205), the IIN, and Northwestern's MRSEC program (NSF DMR-1720139). We gratefully acknowledge support from the Defense Threat Reduction Agency Grant HDTRA1-15-10052/P00001, the National Science Foundation Grants 1936789 and 1844336, the Air Force Research Laboratory Center of Excellence Grant FA8650-15-2-5518, the Bill & Melinda Gates Foundation Grant OPP1217652, the David and Lucile Packard Foundation, and the Camille Dreyfus Teacher-Scholar Program. This project was also supported in part by fellowships awarded to J.M.H. (NDSEG-36373) and K.F.W. (ND-CEN-013-096) through the National Defense Science and Engineering (NDSEG) Fellowship Program, sponsored by the Air Force Research Laboratory, the Office of Naval Research, and the Army Research Office. J.M.H and J.A.P. thank the Ryan Fellowship awarded by Northwestern University. J.A.P. was supported by an NSF Graduate Research Fellowship. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Threat Reduction Agency, or the U.S. Government.

## Competing interests

M.C.J. has a financial interest in Design Pharmaceuticals Inc. and SwiftScale Biologics. M.C.J.'s interests are reviewed and managed by Northwestern University in accordance with their conflict of interest policies. All other authors declare no conflicts of interest. M.P.D. has a financial interest in Glycobia, Inc., Versatope, Inc., Ajuta Therapeutics, Inc. and SwiftScale Biologics. M.P.D.'s interests are reviewed and managed by Cornell University in accordance with their conflict of interest policies.

## ORCIDs:

Jasmine M. Hershewe: 0000-0003-4795-9721
Katherine F. Warfel: 0000-0002-7780-6294
Michael C. Jewett: 0000-0003-2948-6211
Justin A. Peruzzi: 0000-0003-3782-1382
Eric W. Roth: 0000-0002-8295-9834
Neha P. Kamat: 0000-0001-9362-6106
Matthew P. DeLisa: 0000-0003-3226-1566

## Author Contributions

All of the authors designed research; J.M.H., K.F.W., J.A.P., and S.M.I. performed research; C.J.S., J.M.H., J.A.P. and E.W.R. contributed new reagents/analytic tools; J.M.H. and K.F.W. analyzed data; and J.M.H., K.F.W. and M.C.J. wrote the paper. All authors reviewed and edited the paper. M.C.J. provided supervision.

## Data Availability

All data generated or analyzed during this study are included in the manuscript or supplementary information or are available from the corresponding author upon request. Accession codes are listed for applicable genes in the supplementary information. We report no restrictions on data availability.

## References


1. Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 1–20 (2019) doi:10.1038/s41576-019-0186-3.

2. Hershewe, J., Kightlinger, W. & Jewett, M. C. Cell-free systems for accelerating glycoprotein expression and biomanufacturing. *J. Ind. Microbiol. Biotechnol.* (2020) doi:10.1007/s10295-020-02321-4.

3. Rasor, B. *et al.* Toward sustainable, cell-free biomanufacturing. *Curr. Opin. Biotechnol.* **In press**, (2020).

4. Bogart, J. W. *et al.* Cell-free exploration of the natural product chemical space. *ChemBioChem* (2020).

5. Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem* **16**, 2420–2431 (2015).

6. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23**, 150–156 (2005).

7. Kightlinger, W. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun. 2019 101* **10**, 1–13 (2019).

8. Kightlinger, W. *et al.* Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases article. *Nat. Chem. Biol.* **14**, 627–635 (2018).

9. Karim, A. S. *et al.* In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.* (2020) doi:10.1038/s41589-020-0559-0.

10. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.* **30**, 1185–1194 (2012).

11. Kwon, Y. C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **5**, (2015).

12. Liu, D., Zawada, J. & Swartz, J. R. Streamlining Escherichia Coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnol. Prog.* **21**, 460–465 (2005).

13. Borkowski, O. *et al.* Large scale active-learning-guided exploration for in vitro protein production optimization. *Nat. Commun.* **11**, 1872 (2020).

14. Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* (2014) doi:10.1016/j.biochi.2013.11.025.

15. Des Soye, B. J., Gerbasi, V. R., Thomas, P. M., Kelleher, N. L. & Jewett, M. C. A highly productive, one-pot cell-free protein synthesis platform based on genomically recoded Escherichia coli. *Cell Chem. Biol.* (2019) doi:10.1016/j.chembiol.2019.10.008.

16. Contreras-Llano, L. E. *et al.* Holistic engineering of cell-free systems through proteome-reprogramming synthetic circuits. *Nat. Commun.* **11**, 1–10 (2020).

17. Kightlinger, W. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.* **10**, 1–13 (2019).

18. Cai, Q. *et al.* A simplified and robust protocol for immunoglobulin expression in E scherichia coli cell-free protein synthesis systems. *Biotechnol. Prog.* **31**, 823–831 (2015).

19. Bernath, K., Magdassi, S. & Tawfik, D. S. Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. *J. Mol. Biol.* **345**, 1015–1026 (2005).

20. Zawada, J. F. *et al.* Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnol Bioeng* **108**, 1570–1578 (2011).

21. Calhoun, K. A. & Swartz, J. R. An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. *Biotechnol. Prog.* (2005) doi:10.1021/bp050052y.

22. Pardee, K. *et al.* Paper-based synthetic gene networks. *Cell* **159**, 940–954 (2014).

23. Pardee, K. *et al.* Portable, On-Demand Biomolecular Manufacturing. *Cell* **167**, 248-259.e12 (2016).

24. Stark, J. C. *et al.* On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *bioRxiv* (2019) doi:10.1101/681841.

25. Hunt, J. P., Yang, S. O., Wilding, K. M. & Bundy, B. C. The growing impact of lyophilized cell-free protein expression systems. *Bioengineered* **8**, 325–330 (2017).

26. Gregorio, N. E. *et al.* Unlocking Applications of Cell-Free Biotechnology through Enhanced Shelf Life and Productivity of E. coli Extracts. *ACS Synth. Biol.* **9**, 766–778 (2020).

27. Adiga, R. *et al.* Point-of-care production of therapeutic proteins of good-manufacturing-practice quality. *Nat. Biomed. Eng.* **2**, 675–686 (2018).

28. Thavarajah, W. *et al.* Point-of-Use Detection of Environmental Fluoride via a Cell-Free Riboswitch-Based Biosensor. *ACS Synth. Biol.* **9**, 10–18 (2020).

29. Salehi, A. S. M. *et al.* Cell-Free Protein Synthesis Approach to Biosensing hTRβ-Specific Endocrine Disruptors. *Anal. Chem.* **89**, 3395–3401 (2017).

30. Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255–1266 (2016).

31. Liu, X. *et al.* Design of a Transcriptional Biosensor for the Portable, On-Demand Detection of Cyanuric Acid. *ACS Synth. Biol.* acssynbio.9b00348 (2019) doi:10.1021/acssynbio.9b00348.

32. Meyer, A. J. *et al.* Organism engineering for the bioproduction of the triaminotrinitrobenzene (TATB) precursor phloroglucinol (PG). *ACS Synth. Biol.* **8**,

acssynbio.9b00393 (2019).

33. Jung, J. K. *et al.* Cell-free biosensors for rapid detection of water contaminants. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0571-7.

34. Jaroentomeechai, T. *et al.* Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* **9**, 1–11 (2018).

35. Huang, A. *et al.* Biobits$^{TM}$ explorer: A modular synthetic biology education kit. *Sci. Adv.* **4**, (2018).

36. Stark, J. C. *et al.* BioBits$^{TM}$ Bright: A fluorescent synthetic biology education kit. *Sci. Adv.* **4**, (2018).

37. Stark, J. C. *et al.* BioBits Health: Classroom Activities Exploring Engineering, Biology, and Human Health with Fluorescent Readouts. *ACS Synth. Biol.* **8**, 1001–1009 (2019).

38. Martin, R. W. *et al.* Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat. Commun.* (2018) doi:10.1038/s41467-018-03469-5.

39. Oza, J. P. *et al.* Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nat. Commun.* (2015) doi:10.1038/ncomms9168.

40. Silverman, A. D., Akova, U., Alam, K. K., Jewett, M. C. & Lucks, J. B. Design and optimization of a cell-free atrazine biosensor. *ACS Synth. Biol.* **9**, 671–677 (2020).

41. Voyvodic, P. L. *et al.* Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *Nat. Commun.* **10**, 1–8 (2019).

42. Dudley, Q. M., Karim, A. S., Nash, C. J. & Jewett, M. C. Cell-free prototyping of limonene biosynthesis using cell-free protein synthesis. *Metab. Eng.* (2020) doi:10.1016/j.ymben.2020.05.006.

43. Karim, A. S. & Jewett, M. C. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab. Eng.* (2016) doi:10.1016/j.ymben.2016.03.002.

44. Lai, H.-E. *et al.* A GenoChemetic strategy for derivatization of the violacein natural product scaffold. *bioRxiv* (2019).

45. Cappuccio, J. A. *et al.* Cell-free co-expression of functional membrane proteins and apolipoprotein, forming soluble nanolipoprotein particles. *Mol. Cell. Proteomics* **7**, 2246–2253 (2008).

46. Liguori, L., Marques, B. & Lenormand, J. L. A bacterial cell-free expression system to produce membrane proteins and proteoliposomes: From cDNA to functional assay. *Current Protocols in Protein Science* (2008) doi:10.1002/0471140864.ps0522s54.

47. Matthies, D. *et al.* Cell-free expression and assembly of ATP synthase. *J. Mol. Biol.*

(2011) doi:10.1016/j.jmb.2011.08.055.

48. Sachse, R., Dondapati, S. K., Fenz, S. F., Schmidt, T. & Kubick, S. Membrane protein synthesis in cell-free systems: From bio-mimetic systems to bio-membranes. *FEBS Letters* (2014) doi:10.1016/j.febslet.2014.06.007.

49. Schwarz, F. *et al.* Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology* (2011) doi:10.1093/glycob/cwq130.

50. Schoborg, J. A. *et al.* A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnology and Bioengineering* (2017) doi:10.1002/bit.26502.

51. Tai, P. C., Tian, G., Xu, H., Lian, J. P. & Jack, N. Y. In vitro protein translocation into Escherichia coli inverted membrane vesicles. *Methods Cell Biol.* **34**, 167–187 (1991).

52. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* (2001) doi:10.1006/jmbi.2000.4315.

53. Aebersold, R. *et al.* How many human proteoforms are there? *Nature Chemical Biology* vol. 14 206–214 (2018).

54. Wuu, J. J. & Swartz, J. R. High yield cell-free production of integral membrane proteins without refolding or detergents. *Biochimica et Biophysica Acta - Biomembranes* vol. 1778 1237–1250 (2008).

55. Goerke, A. R. & Swartz, J. R. Development of cell-free protein synthesis platforms for disulfide bonded proteins. *Biotechnol. Bioeng.* **99**, 351–367 (2008).

56. Goerke, A. R. & Swartz, J. R. High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol. Bioeng.* **102**, 400–416 (2009).

57. Altendorf, K. H. & Staehelin, L. A. Orientation of membrane vesicles from Escherichia coli as detected by freeze-cleave electron microscopy. *J. Bacteriol.* **117**, 888–899 (1974).

58. Hertzberg, E. L. & Hinkle, P. C. Oxidative phosphorylation and proton translocation in membrane vesicles prepared from Escherichia coli. *Biochem. Biophys. Res. Commun.* (1974) doi:10.1016/0006-291X(74)90908-5.

59. Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J. & Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* **4**, (2008).

60. Katzen, F. & Kudlicki, W. Efficient generation of insect-based cell-free translation extracts active in glycosylation and signal sequence processing. *J. Biotechnol.* **125**, 194–197 (2006).

61. Zemella, A. *et al.* Cell-free protein synthesis as a novel tool for directed glycoengineering of active erythropoietin. *Sci. Rep.* **8**, 1–12 (2018).

62. Gurramkonda, C. *et al.* Improving the recombinant human erythropoietin glycosylation using microsome supplementation in CHO cell-free system. *Biotechnol. Bioeng.* **115**, 1253–1264 (2018).

63. Buntru, M., Vogel, S., Stoff, K., Spiegel, H. & Schillberg, S. A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates. *Biotechnol. Bioeng.* **112**, 867–878 (2015).

64. Brödel, A. K. *et al.* IRES-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems. *PLoS One* **8**, (2013).

65. Sachse, R. *et al.* Synthesis of membrane proteins in eukaryotic cell-free systems. *Eng. Life Sci.* **13**, 39–48 (2013).

66. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).

67. Kuriakose, A., Chirmule, N. & Nair, P. Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications. *J. Immunol. Res.* **2016**, (2016).

68. Parodi, A. J. Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J.* **13**, 1–13 (2000).

69. Zheng, K., Bantog, C. & Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* **3**, 568–576 (2011).

70. Kightlinger, W., Warfel, K. F., DeLisa, M. P. & Jewett, M. C. Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth. Biol.* **40**, 7 (2020).

71. DeLisa, M. *et al.* Cell-free synthetic glycobiology: designing and engineering glycomolecules outside of living cells. *Front. Chem.* **8**, 645 (2020).

72. Natarajan, A. *et al.* Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. *Nat. Chem. Biol.* (2020) doi:10.1038/s41589-020-0595-9.

73. Underwood, K. A., Swartz, J. R. & Puglisi, J. D. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng.* (2005) doi:10.1002/bit.20529.

74. Ramakrishnan, V. Ribosome Structure and the Mechanism of Translation. *Cell* **108**, 557–572 (2002).

75. Monguió-Tortajada, M., Gálvez-Montón, C., Bayes-Genis, A., Roura, S. & Borràs, F. E. Extracellular vesicle isolation methods: rising impact of size-exclusion chromatography. *Cell. Mol. Life Sci.* **76**, 2369–2382 (2019).

76. Mol, E. A., Goumans, M. J., Doevendans, P. A., Sluijter, J. P. G. & Vader, P. Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation. *Nanomedicine Nanotechnology, Biol. Med.* **13**, 2061–2065 (2017).

77. Böing, A. N. *et al.* Single-step isolation of extracellular vesicles by size-exclusion chromatography. *J. Extracell. Vesicles* **3**, (2014).

78. Cole, S. D., Miklos, A. E., Chiao, A. C., Sun, Z. Z. & Lux, M. W. Methodologies for preparation of prokaryotic extracts for cell-free expression systems. *Synth. Syst. Biotechnol.* **5**, 252–267 (2020).

79. Huang, C., Quinn, D., Sadovsky, Y., Suresh, S. & Hsia, K. J. Formation and size distribution of self-Assembled vesicles. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2910–2915 (2017).

80. Young, N. M. *et al.* Structure of the N-linked glycan present on multiple glycoproteins in the gram-negative bacterium, Campylobacter jejuni. *J. Biol. Chem.* (2002) doi:10.1074/jbc.M206114200.

81. Faridmoayer, A. *et al.* Extreme substrate promiscuity of the Neisseria oligosaccharyl transferase involved in protein O-glycosylation. *J. Biol. Chem.* **283**, 34596–34604 (2008).

82. Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S. & Feldman, M. F. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J. Bacteriol.* **189**, 8088–8098 (2007).

83. Pan, C. *et al.* Biosynthesis of conjugate vaccines using an O-linked glycosylation system. *MBio* **7**, (2016).

84. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (80-. ).* **298**, 1790–1793 (2002).

85. Ma, Z. *et al.* Glycoconjugate Vaccine Containing Escherichia coli O157:H7 O-Antigen Linked with Maltose-Binding Protein Elicits Humoral and Cellular Responses. *PLoS One* **9**, e105215 (2014).

86. Palmu, A. A. *et al.* Effectiveness of the ten-valent pneumococcal Haemophilus influenzae protein D conjugate vaccine (PHiD-CV10) against invasive pneumococcal disease: A cluster randomised trial. *Lancet* **381**, 214–222 (2013).

87. Jewett, M. C. & Swartz, J. R. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* **86**, 19–26 (2004).

88. Cole, S. D. *et al.* Quantification of Interlaboratory Cell-Free Protein Synthesis Variability. *ACS Synth. Biol.* **8**, 2080–2091 (2019).

89. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8**, 434–436 (2012).

90. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976 (2006).

91. Guarino, C. & Delisa, M. P. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* **22**, 596–601 (2012).

92. Fishov, I. & Woldringh, C. L. Visualization of membrane domains in Escherichia coli. *Mol. Microbiol.* (1999) doi:10.1046/j.1365-2958.1999.01425.x.

93. McBroom, A. J., Johnson, A. P., Vemulapalli, S. & Kuehn, M. J. Outer membrane vesicle production by Escherichia coli is independent of membrane instability. *J. Bacteriol.* **188**, 5385–5392 (2006).

## Figures and Figure Legends

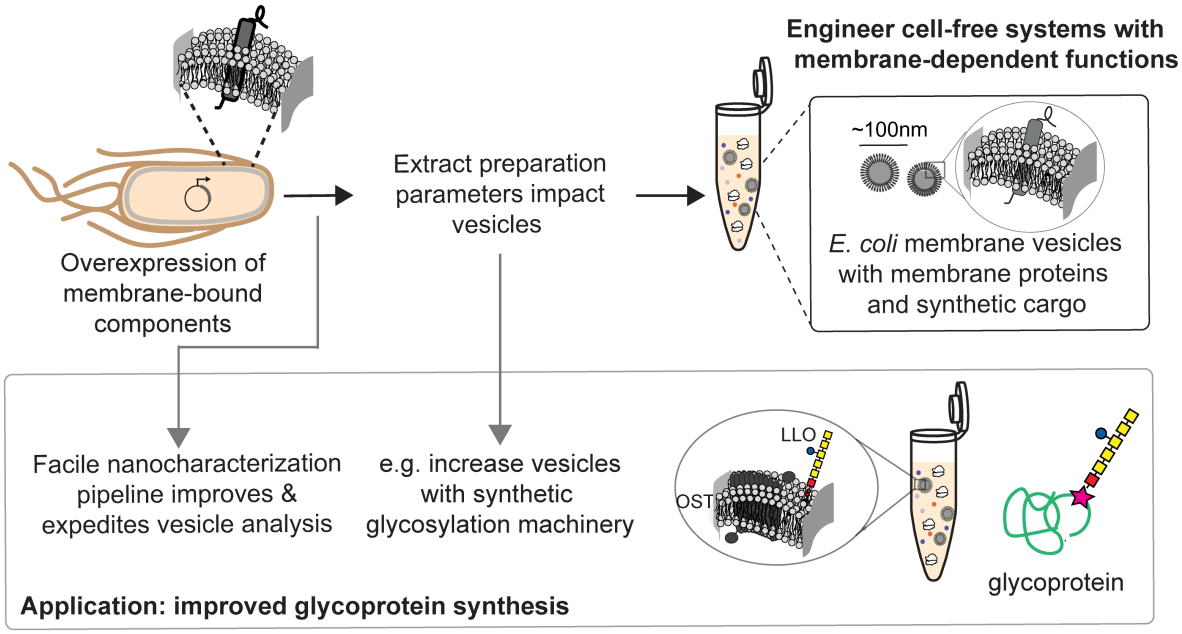



**Figure 1. Overview schematic of engineering CFE systems with cell-derived membrane-dependent functions.** Membrane-bound cargo expressed in living *E. coli* is carried through into CFE extracts via membrane vesicles. The extract preparation method used to prepare CFE extracts impacts sizes and concentrations of vesicles, and their associated cargo. Here, we developed a facile nanocharacterization pipeline to better understand and characterize the impacts of extract preparation methods on vesicle profiles and their associated cargo. We then apply our findings to improve cell-free glycoprotein synthesis, which is a promising platform for on-demand vaccine development. By increasing concentrations of vesicles and membrane-bound glycosylation machinery (OST and LLO), we overcome limitations in cell-free glycoprotein synthesis and significantly increase glycoprotein titers.

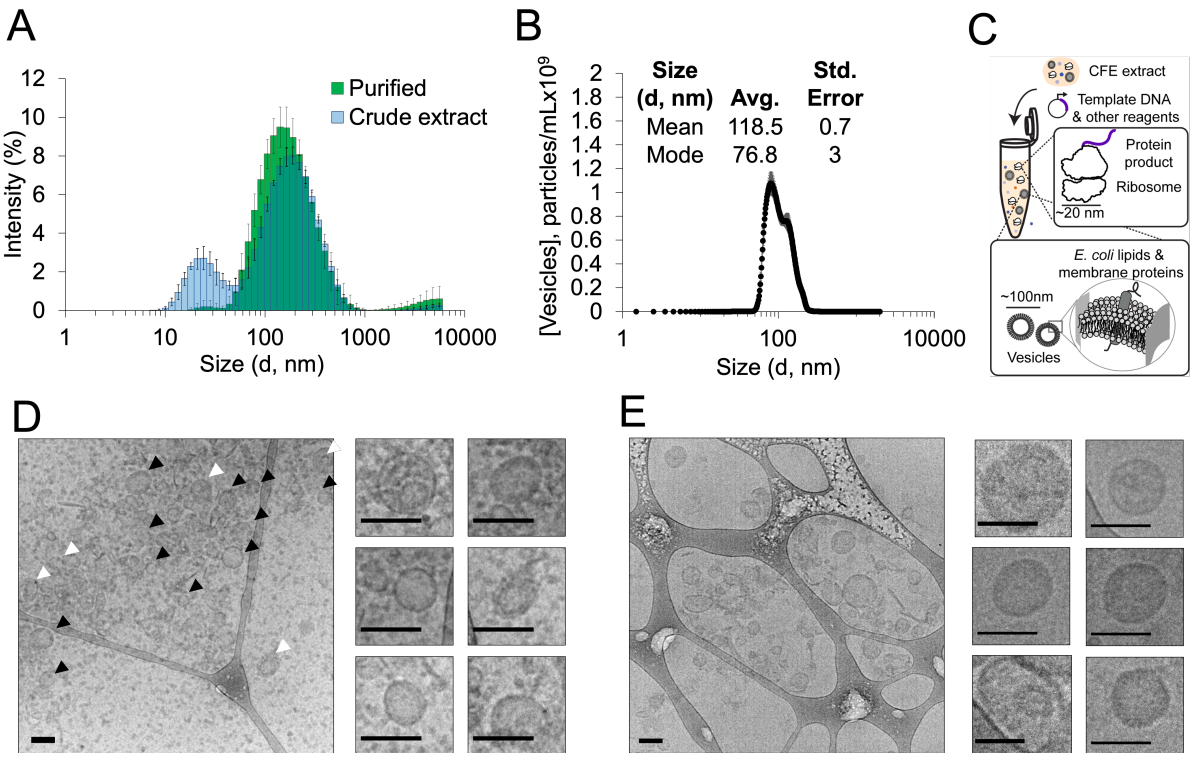



**Figure 2. Characterization of membrane vesicles in crude CFE extracts. (A)** DLS analysis of crude extracts and SEC purified vesicles. Error bars represent the standard deviation within triplicate analysis of three independently-prepared extracts. For purified vesicles, error bars represent the standard deviation of triplicate analysis of the most concentrated vesicle elution fraction. **(B)** NTA of purified vesicles collected from SEC. Mean and mode diameters observed in the particle size distribution are listed in the inset. **(C)** Illustration of particles detected in crude CFE extracts. **(D)** CryoEM micrographs of crude extracts. Black arrows indicate vesicles with apparent unilamellar morphology. White arrows indicate nested or multilamellar morphologies. Cropped images indicate representative vesicles. Scale bars bars are 100 nm. Uncropped images are available in **Supplementary Fig. 3** and numbered with the corresponding cropped vesicles. **(E)** CryoEM micrographs of SEC purified vesicles. Cropped images indicate representative purified vesicle particles. Scale bars are 100 nm. Uncropped images are available in **Supplementary Fig. 3** and numbered with the corresponding cropped vesicles.

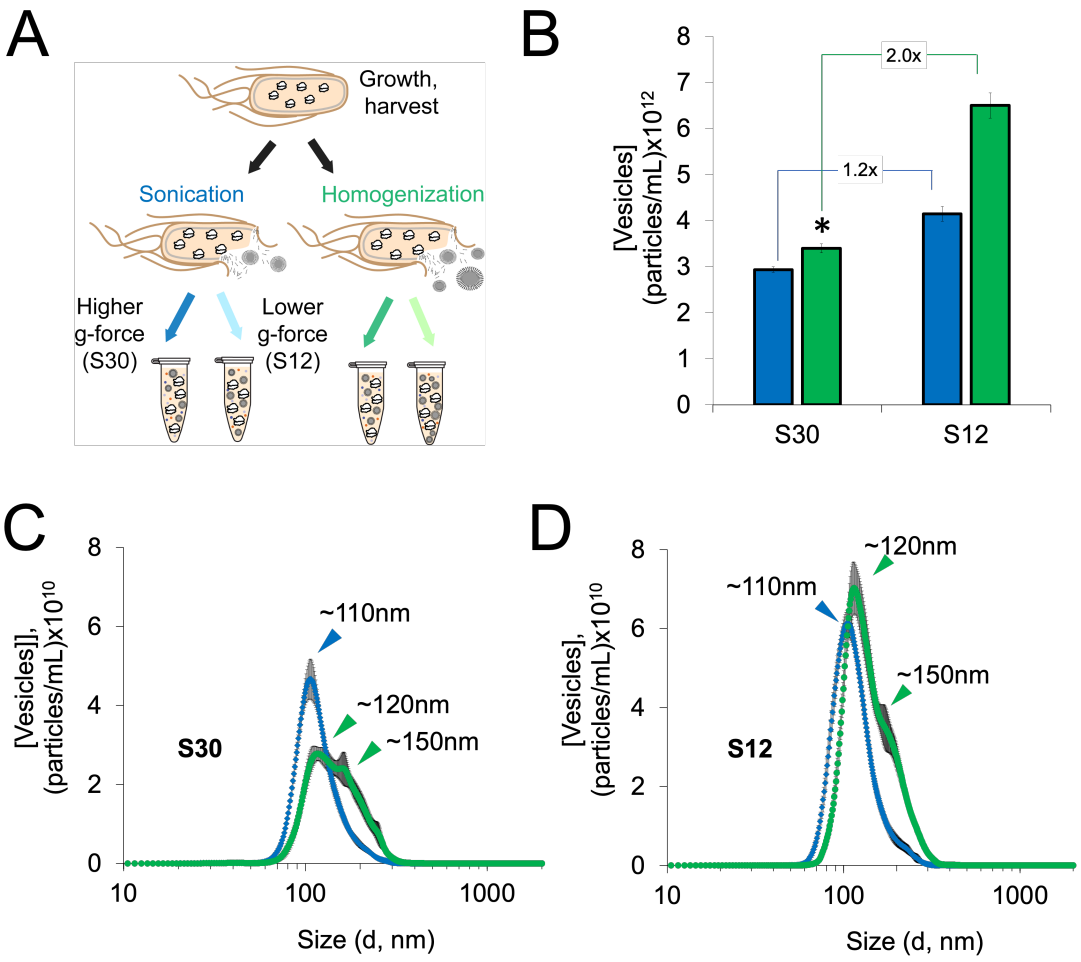



**Figure 3. Extract processing impacts vesicle size distributions and concentrations. (A)** Illustration of extract processing conditions. Extracts were prepared in triplicate for each condition shown. **(B)** Nanoparticle Tracking Analysis (NTA) concentration analysis of vesicles in sonicated (blue) and homogenized (green) extracts. Asterisk indicates base case conditions for extract preparation. **(C)** NTA particle size distribution of vesicles in sonicated (blue) and homogenized (green) S30 extracts. **(D)** NTA particle size distribution of sonicated (blue) and homogenized (green) S12 extracts. Error bars represent the standard error of the mean for NTA measurements of three independent extracts.

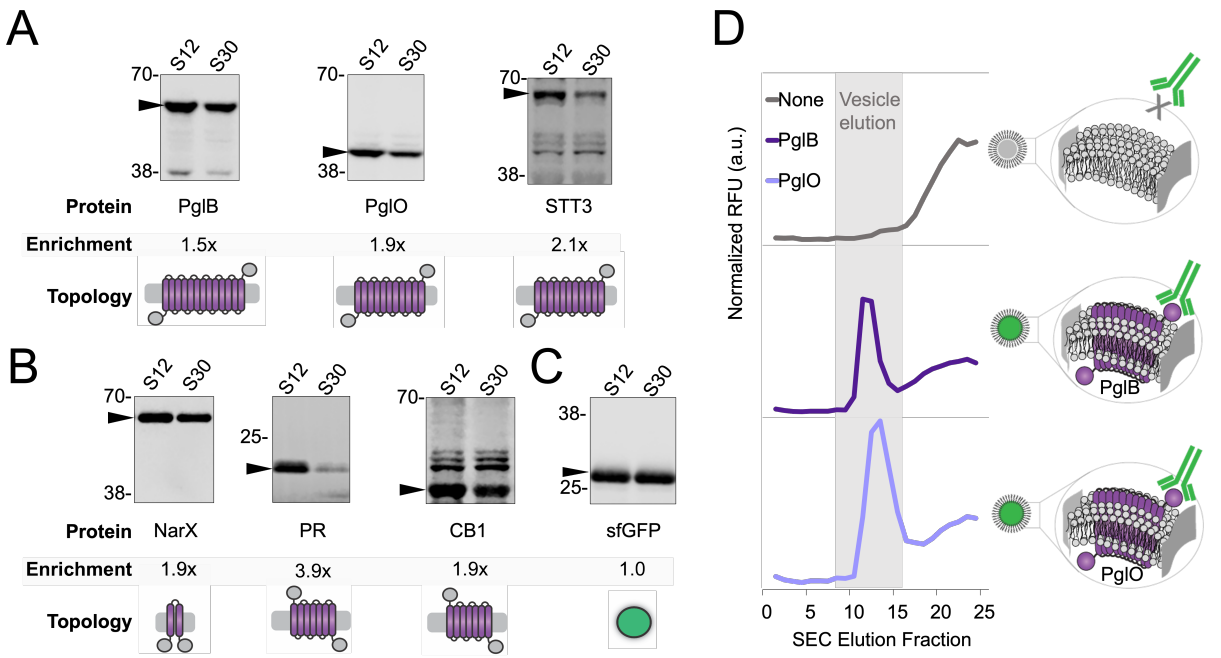



**Figure 4. Heterologous membrane-bound cargo can be controllably enriched via membrane vesicles.** Enrichment of heterologous membrane proteins in S12 and S30 extracts was quantitated using α-FLAG Western blots against the heterologous proteins for **(A)** glycosylation enzymes and **(B)** signal transduction proteins. **(C)** Cytosolic sfGFP control with no transmembrane helices. On each Western blot, left lanes are S12 extracts and right lanes are S30 extracts. Black arrow indicates the membrane protein of interest. Molecular weight (kDa) from protein ladder standards are indicated to the left of each blot. Protein names and enrichment ratio of bands (S12/S30) are shown directly below each blot. All blots are representative of 3 independent experiments. Cartoons depict the transmembrane topology for each protein. See **Supplementary Table 1** for taxonomical origin, transmembrane topology, functions(s), theoretical size, and UniProt ID. **(D)** Fluorescence chromatograms of SEC analysis of extracts probed with a fluorescent α-FLAG antibody. Strains used to prepare extracts were enriched with no membrane protein (gray trace), PglB (dark purple trace), or PglO (light purple trace). Characteristic vesicle elution fraction from 3 independent experiments is highlighted in gray.

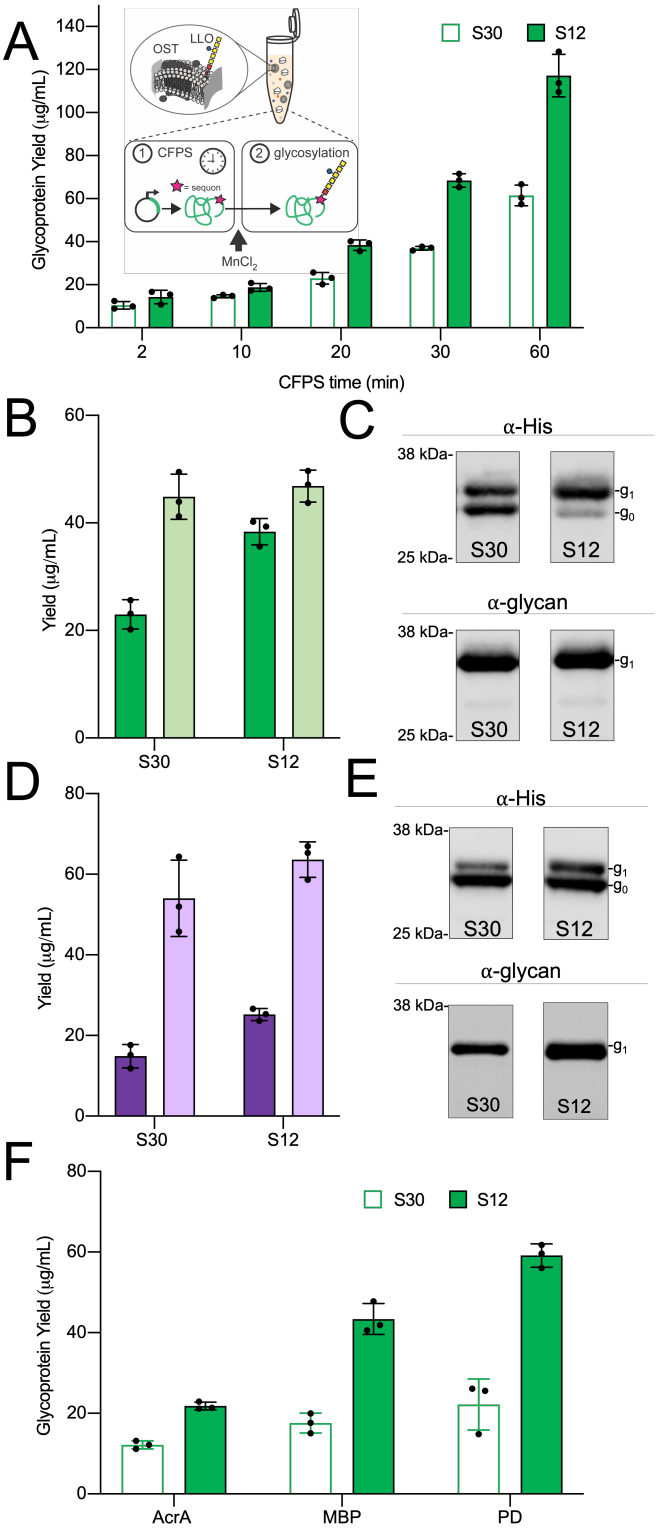



***Figure 5.*** **Increasing vesicle concentrations improves cell-free glycoprotein synthesis (CFGpS) for *N*- and *O*-linked glycosylation systems.** For panels **A-E** a standard curve correlating protein yields derived from $^{14}$C-Leucine counting and sfGFP florescence was used to measure protein concentrations. Quantitative Western blotting was used to measure fraction of

glycosylated protein. For panel **F**, protein concentrations were measured using $^{14}$C-Leucine incorporation. Fraction of glycosylated protein was measured using autoradiography. **(A)** sfGFP glycoprotein yields of CFGpS reactions charged with S12 or S30 extracts enriched with PglB and *C. jejuni* LLO. Error bars represent standard deviation of 3 independent CFGpS reactions, each run with an independent extract. **(Inset)** Schematic of 2-phase CFGpS reactions. **(B)** Glycosylated (dark green) and total (light green) sfGFP yields of *N*-linked CFGpS reactions with 20-min CFPS times. Error bars represent standard deviation of 3 independent reactions. **(C)** Western blots of acceptor proteins from representative reactions in **(B). (D)** Glycosylated (dark purple) and total (light light) sfGFP yields from *O*-linked CFGpS reactions with 20 min CFPS times. **(E)** Western blots of acceptor proteins from representative reactions in **(D)**. **(F)** Glycoprotein yields of AcrA, MBP, and PD produced in CFGpS reactions charged with S12 or S30 extracts enriched with PglB and *C. jejuni* LLO. Error bars represent standard deviation of 3 independent CFGpS reactions, each run with an independent extract. Glycoprotein yields of AcrA, which contains 2 internal glycosylation sites, include singly and doubly glycosylated protein. Error bars represent standard deviation of 3 independent reactions.

# Supplementary Information for

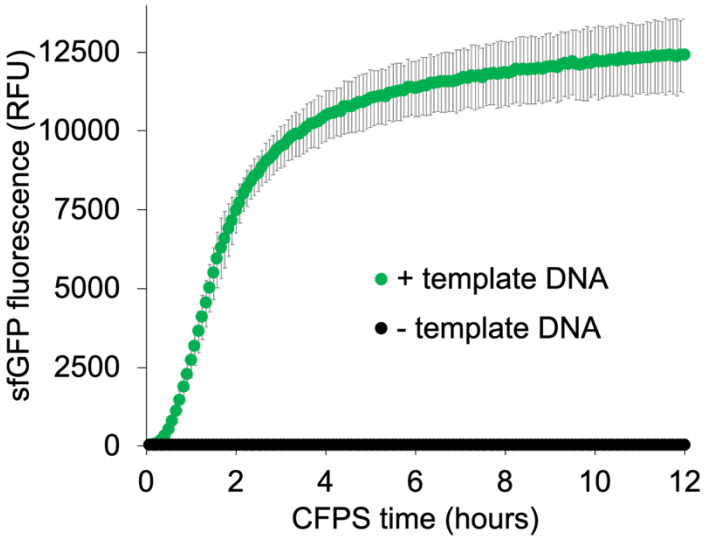
## Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles

Jasmine M. Hershewe[a,b,c,1], Katherine F. Warfel[a,b,c,1], Shaelyn M. Iyer[a], Justin A. Peruzzi[a,b,c], Claretta J. Sullivan[d], Eric W. Roth[e], Matthew P. DeLisa[f,g,h], Neha P. Kamat[b,c,i], and Michael C. Jewett[a,b,c,j,k 2]

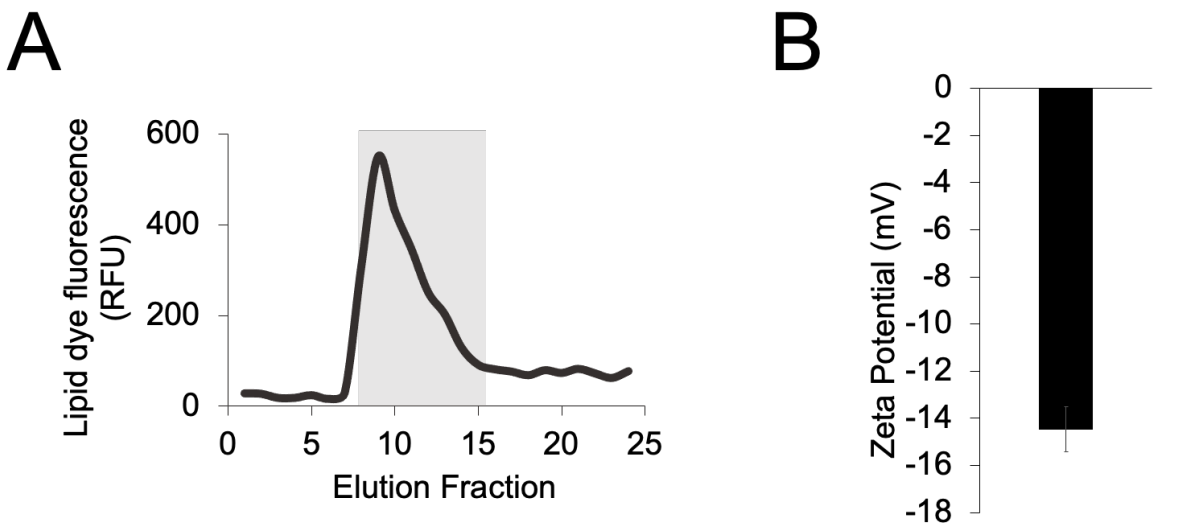
Affiliations

[a]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, IL 60208, USA
[b]Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, USA
[c]Center for Synthetic Biology, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, IL 60208, USA
[d]Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
[e]Northwestern University Atomic and Nanoscale Characterization and Experimentation (NU*ANCE*) Center, 2145 Sheridan Road, Tech Institute A/B Wing A173, Evanston, IL 60208, USA
[f]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853, USA
[g]Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA
[h]Biomedical and Biological Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA
[i]Department of Biomedical Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E310, Evanston, IL 60208, USA
[j]Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North Saint Clair Street, Suite 1200, Chicago, IL 60611, USA
[k]Simpson Querrey Institute, Northwestern University, 303 East Superior Street, Suite 11-131, Chicago, IL 60611, USA

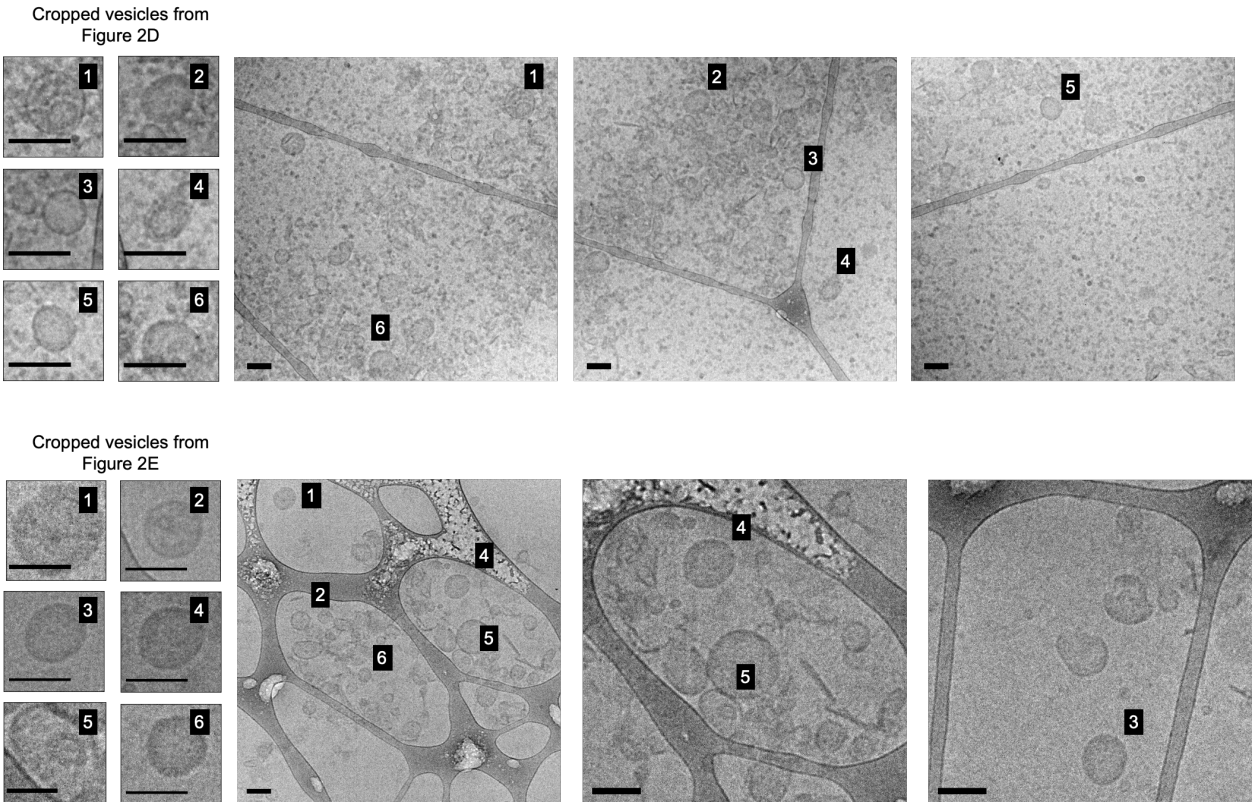
[1]Authors contributed equally
[2]To whom correspondence should be addressed
*Michael C. Jewett
**Phone:** 1 847-497-5007
**Email:** m-jewett@northwestern.edu



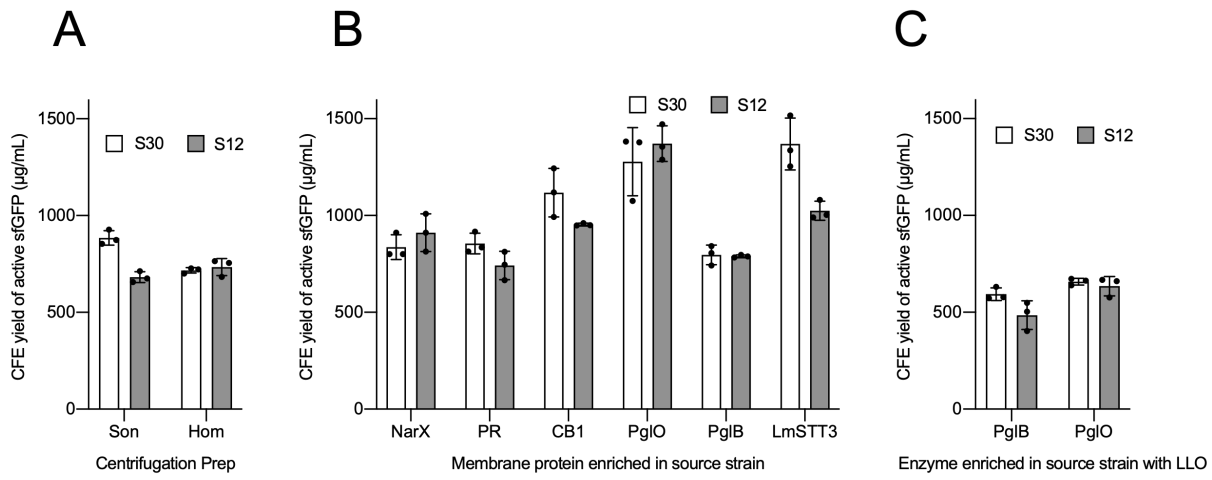

**Supplementary Figure 1**. Production of superfolder green fluorescent protein (sfGFP) in standard cell-free expression (CFE) reaction conditions. Protein synthesis proceeds when template DNA is present, and no fluorescence is observed when template DNA is omitted. Error bars represent standard deviations of three independent CFE reactions.



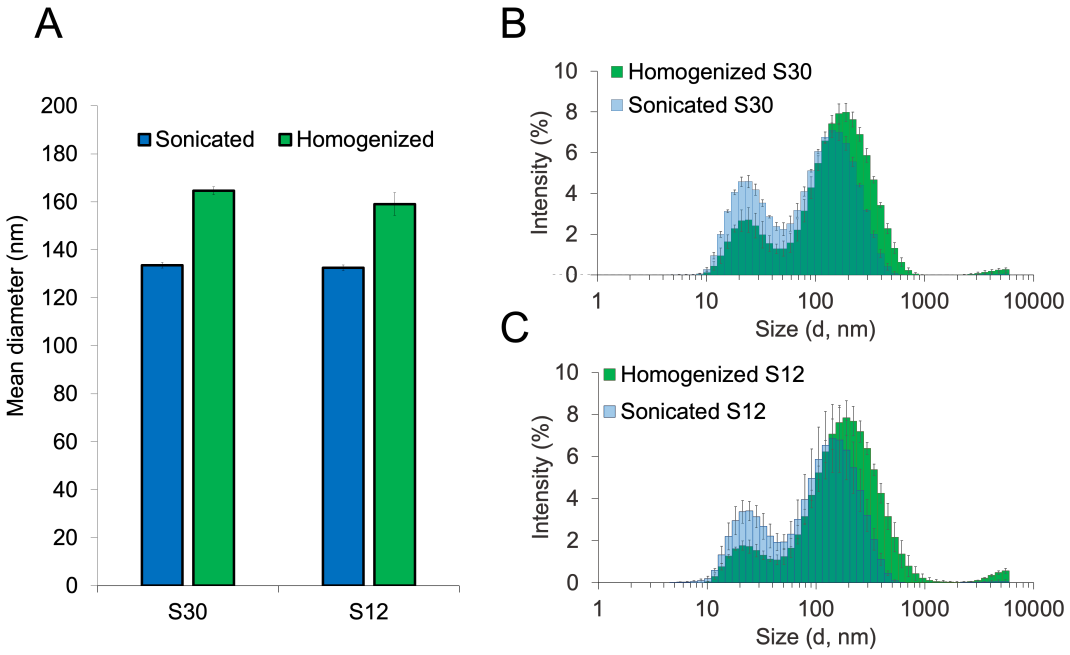

**Supplementary Figure 2**. Purification and characterization of membrane vesicles from CFE extracts. **(A)** SEC chromatogram of extracts probed with FM 4-64 lipid dye. The gray segment indicates the characteristic vesicle elution fraction. **(B)** Zeta potential analysis of purified vesicles in PBS. Error bars represent standard deviation of triplicate measurements of purified vesicles.



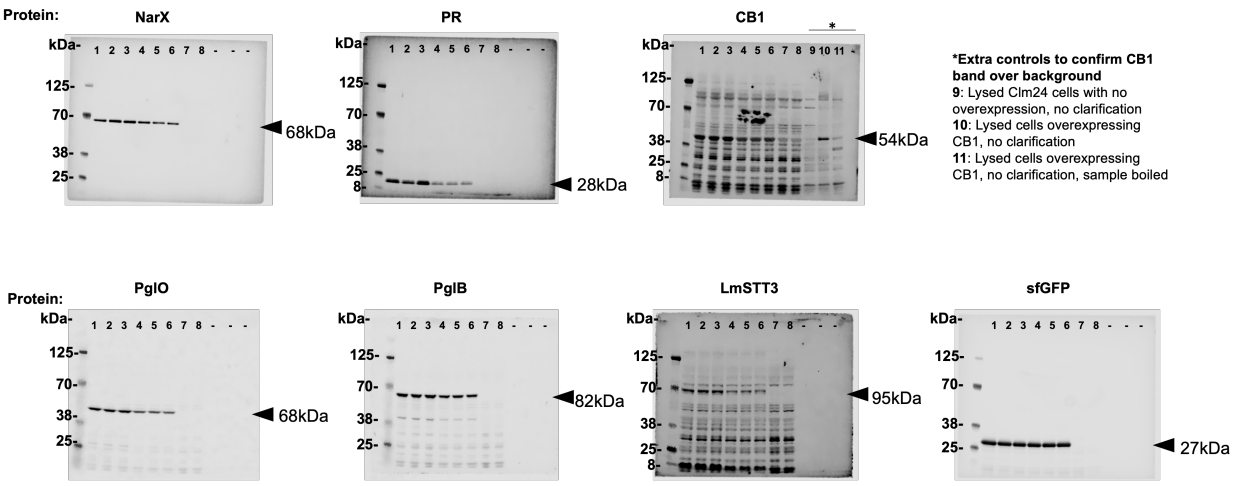

**Supplementary Figure 3.** Cryo-EM analysis of vesicles in crude extracts (top) and after SEC purification (bottom). The numbering of vesicles in uncropped images corresponds with the vesicle shown in each cropped image in **Figure 2** of the main text.




**Supplementary Figure 4.** CFE productivities of all S30 and S12 extracts characterized in this study**.** Reactions were run for 20 hours at 30 °C under standard conditions. **(A)** Extracts made from 'blank' chassis strains with no overexpressed components. These extracts are characterized in **Figures 1** and **2** of the main text**. (B)** Extracts enriched with membrane proteins, characterized in **Figure 3** of the main text**. (C)** CFGpS extracts with enrichment of *C. jejuni* LLO and the denoted enzyme in the strain. CFGpS extracts are characterized in **Figure 4** of the main text. Error bars represent standard deviations of triplicate independent reactions.



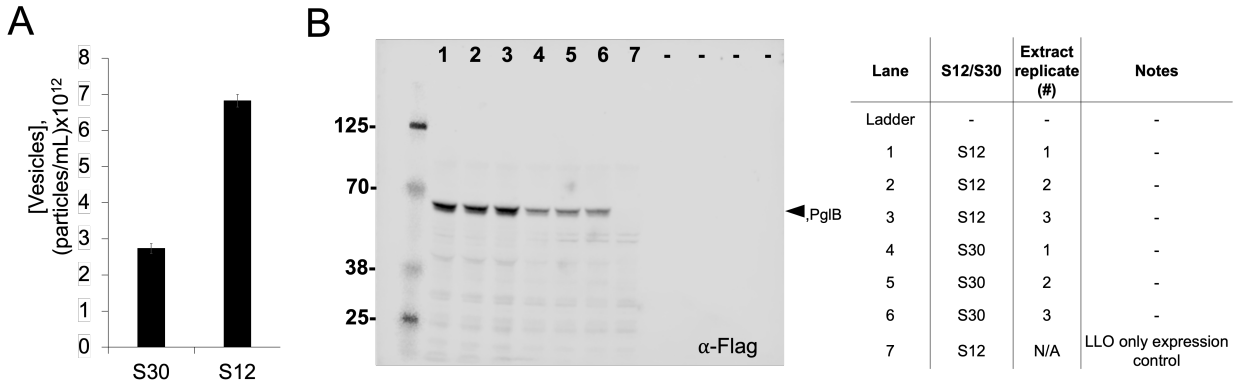

**Supplementary Figure 5**. Additional light scattering characterization of extracts presented in **Figure 3** of the main text. **(A)** Mean vesicle diameters, determined by NTA. Error bars represent the standard error of the mean for NTA measurements of three independent extracts. DLS analysis of **(B)** S30 extracts and **(C)** S12 extracts. Spectra corroborate larger, right-shifted vesicle peaks in homogenized extracts in both cases. Consistent with NTA particle counting, the relative peak heights of ~20 nm peak (ribosomes/small cellular complexes) to vesicle peak indicates that homogenized extracts contain higher concentrations of vesicles than sonicated extracts for each given prep method. Error bars represent the standard deviation of measurements of three independent extracts.




| Lane | S12/S30 | Extract replicate (#) | Notes |
|---|---|---|---|
| Ladder | - | - | - |
| 1 | S12 | 1 | - |
| 2 | S12 | 2 | - |
| 3 | S12 | 3 | - |
| 4 | S30 | 1 | - |
| 5 | S30 | 2 | - |
| 6 | S30 | 3 | - |
| 7 | S12 | N/A | No overexpression, strain background control |
| 8 | S30 | N/A | No overexpression, strain background control |

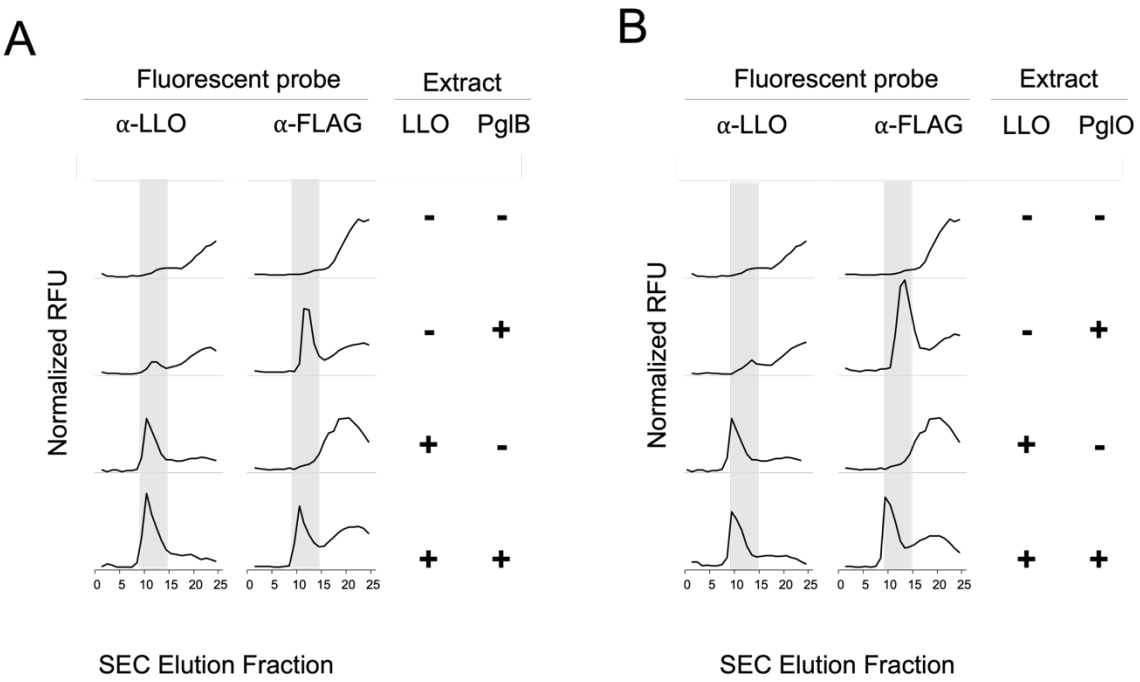
| Protein | S12 avg densitometry signal | S12 std dev densitometry signal | S30 avg densitometry signal | S30 std dev densitometry signal | S12/S30 densitometry signal | S12/S30 error densitometry signal | S12/S30 ratio rounded for Figure |
|---|---|---|---|---|---|---|---|
| NarX | 2.76 | 0.70 | 1.44 | 0.37 | 1.91 | 0.69 | 1.9 |
| PR | 0.83 | 0.22 | 0.21 | 0.03 | 3.88 | 1.16 | 3.9 |
| CB1 | 0.25 | 0.01 | 0.13 | 0.01 | 1.88 | 0.11 | 1.9 |
| PglO | 8.71 | 0.36 | 4.66 | 0.37 | 1.87 | 0.17 | 1.9 |
| PglB | 5.23 | 0.10 | 3.47 | 0.18 | 1.51 | 0.08 | 1.5 |
| LmSTT3 | 0.16 | 0.01 | 0.08 | 0.01 | 2.14 | 0.12 | 2.1 |
| sfGFP | 16.87 | 0.72 | 16.90 | 0.26 | 1.00 | 0.05 | 1.0 |

**Supplementary Figure 6**. Western blot analysis of membrane protein enrichment in S30 and S12 extracts. Uncropped α-FLAG blots against each of the indicated recombinant proteins. The theoretical mass of each recombinant protein is listed next to black arrows, indicating the corresponding band. We observe the well-documented effect that membrane proteins run anomalously on SDS-PAGE, running 'light' with respect to the protein ladder standard. A lane key is presented below. Note that extra controls (indicated to the right of blot) were needed for the CB1 blot to confirm the presence of the protein. Averages and standard deviations of densitometry signals used to determine S12/S30 enrichment in **Figure 4** are also included and are reported rounded to the nearest hundredths place.

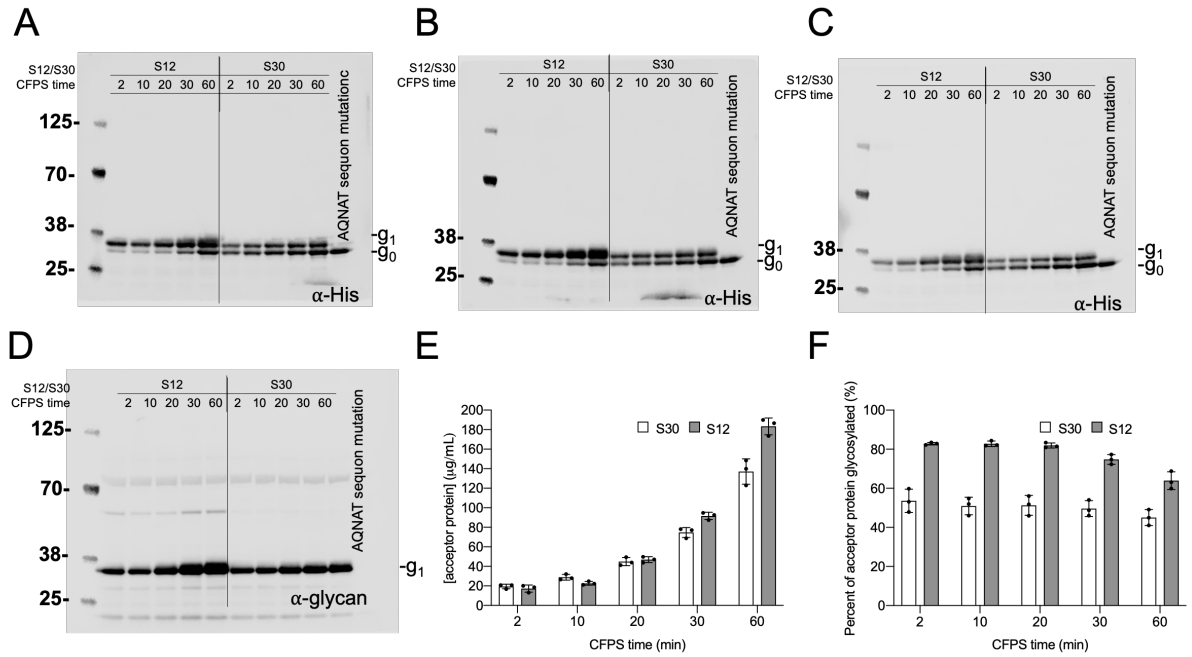



| Lane | S12/S30 | Extract replicate (#) | Notes |
|---|---|---|---|
| Ladder | - | - | - |
| 1 | S12 | 1 | - |
| 2 | S12 | 2 | - |
| 3 | S12 | 3 | - |
| 4 | S30 | 1 | - |
| 5 | S30 | 2 | - |
| 6 | S30 | 3 | - |
| 7 | S12 | N/A | LLO only expression control |

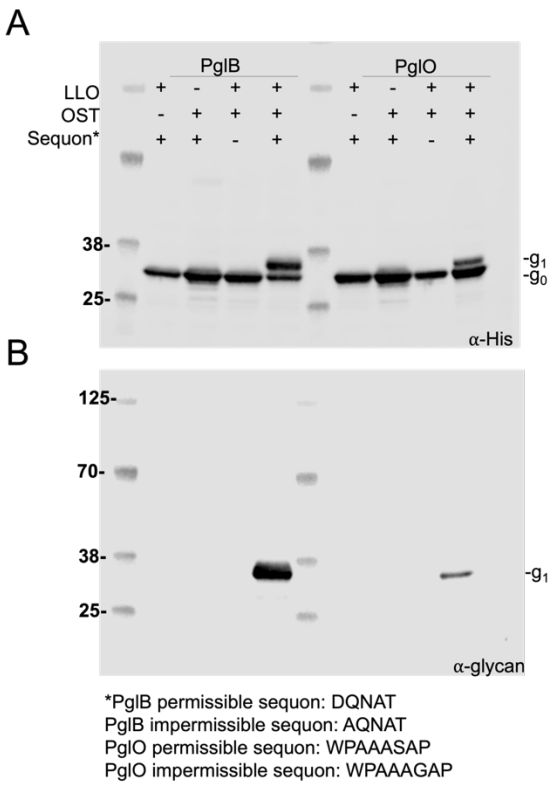
**Supplementary Figure 7**. Characterization of vesicle and PglB enrichment in CFGpS extracts. **(A)** Concentrations of vesicles in extracts enriched with PglB and *C. jejuni* LLO, measured via NTA. Error bars represent standard deviation of measurements of three independent extracts. **(B)** α-FLAG Western blot against PglB in an extract enriched with both PglB and LLO. Corresponding lane key is to the right of the gel.



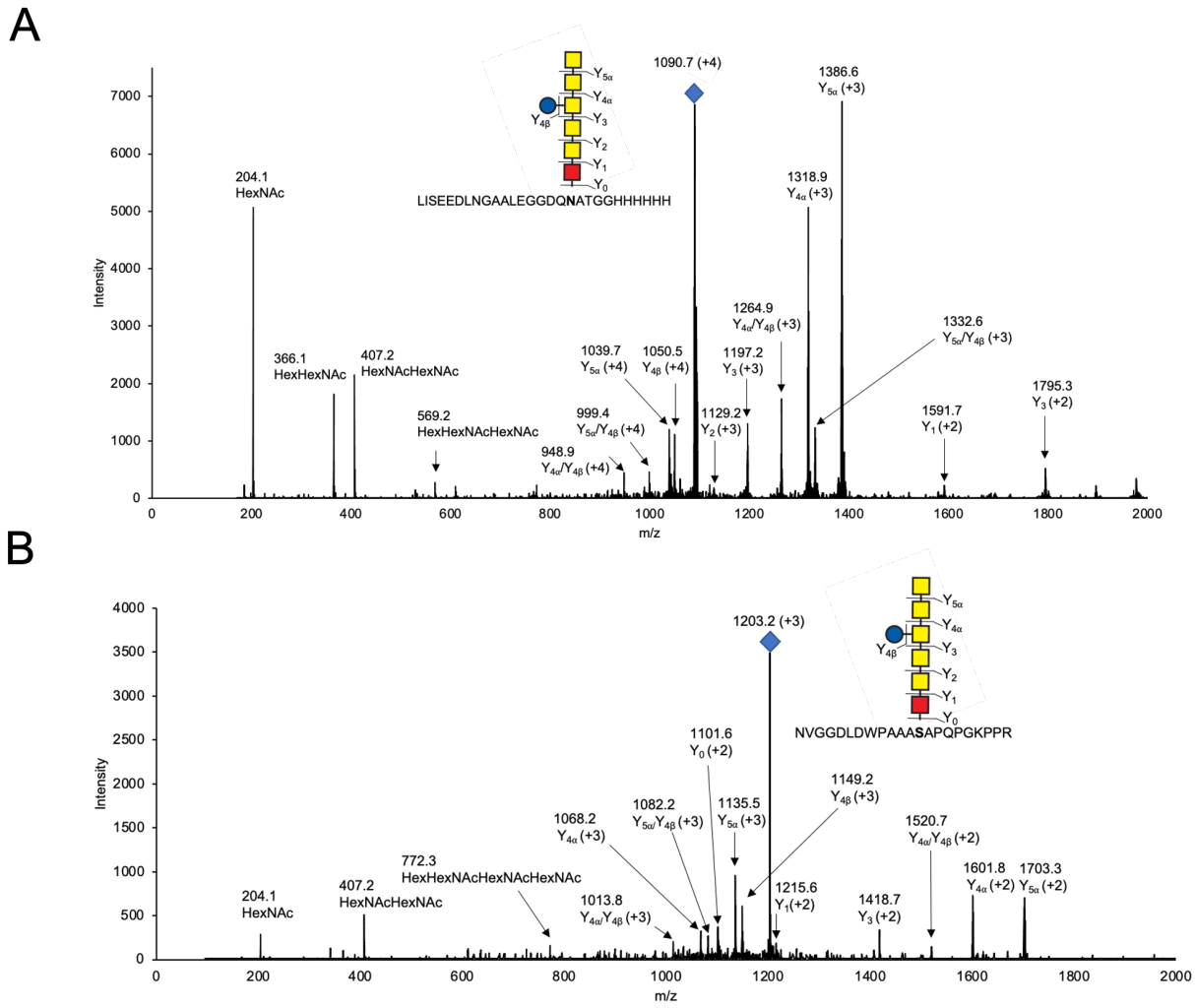

**Supplementary Figure 8**. Fluorescent probes confirm that glycosylation components are embedded in membrane vesicles in CFE extracts. Fluorescence SEC chromatograms of vesicles probed with α-LLO and α-OST reagents are presented. Vesicle elution fraction is highlighted in gray. Analysis of extracts and controls with the *N*-linked PglB OST are presented in **(A)** and for the *O*-linked PglO OST in **(B)**. A low amount of nonspecific binding of the α-LLO SBA lectin is observed and serves as a signal baseline for the LLO-containing samples.




**Supplementary Figure 9**. Characterization of *N*-linked glycosylation in S12 and S30 CFGpS extracts. Triplicate α-His Western blots against CFGpS-derived acceptor proteins are shown in **(A)-(C)** corresponding with extract replicates 1-3. Western blots in **(A)-(C)** were used to calculate glycoprotein yields in **Figure 4** of the main text. **(D)** α-glycan blot of the corresponding reactions in **(A).** $g_o$ denotes aglycosylated acceptor protein and $g_1$ indicates glycoprotein. **(E)** Total acceptor protein produced as determined by sfGFP fluorescence converted to concentration using a $^{14}$C-leucine derived standard curve, and **(F)** percent of acceptor protein converted to glycoprotein at each condition. Error bars represent standard deviation of three independent reactions.

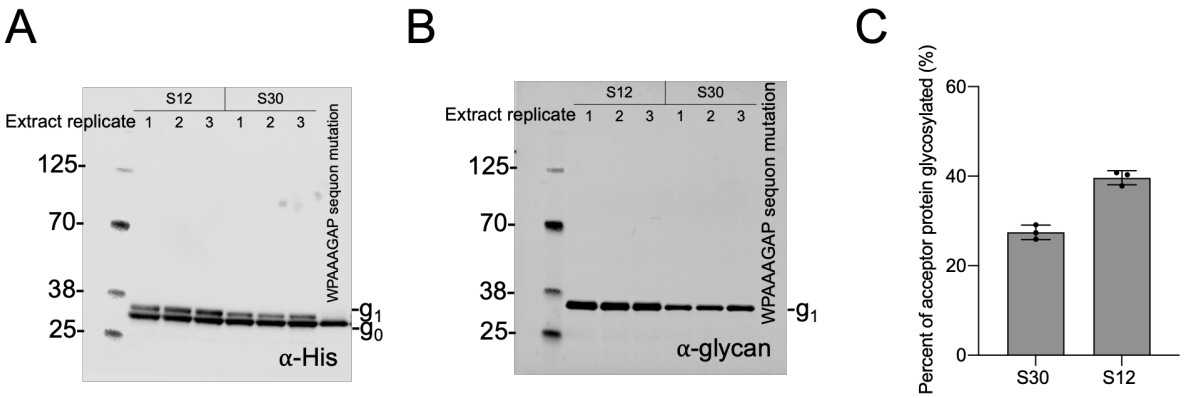



**Supplementary Figure 10.** Residue-specific glycosylation of acceptor proteins with permissible sequons for PglB and PglO. Generally, glycosylation preferences of PglO are less understood then those for PglB. PglB samples are used as blotting references for positive and negative controls, as the sequon specificities for positive and negative sequons are well-characterized. **(A)** α-His blot of CFGpS reactions run with a 5-minute CFPS time. Glycosylated band is only present when all glycosylation components and a permissible sequon are present. $g_o$ denotes aglycosylated acceptor protein and $g_1$ indicates glycoprotein. **(B)** α-glycan blot that corresponds to data from (**A**), showing a signal only for $g_1$ when all glycosylation components are present. Gels are representative of three independent reactions.

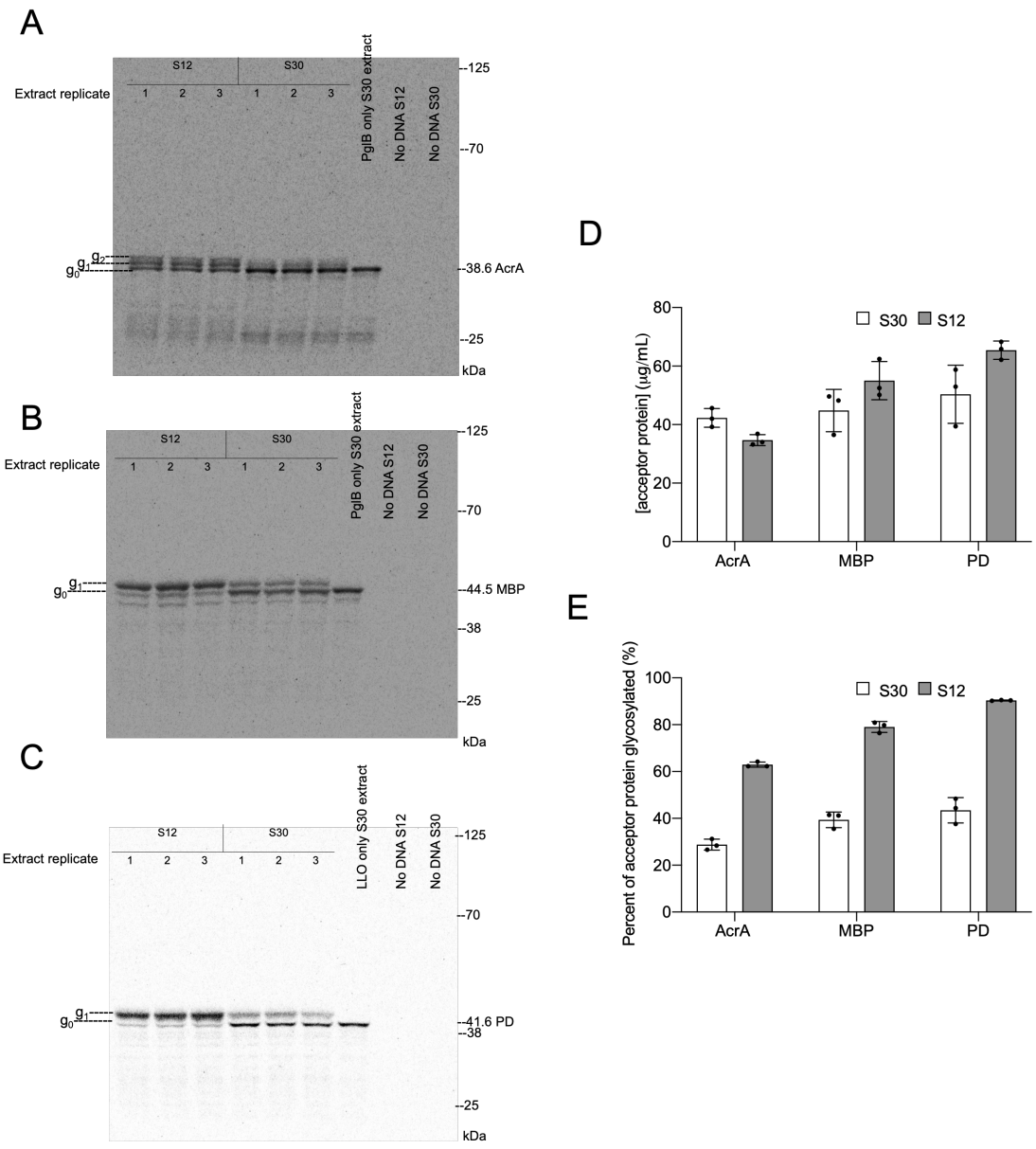



**Supplementary Figure 11.** LC-MS/MS of trypsin digested glycopeptides**.** LC-MS/MS was performed with a Bruker Elute UPLC coupled to an Impact-II UHR TOF Mass Spectrometer. (**A**) A quadruply-charged precursor ion (denoted with a blue diamond) was identified as the glycopeptide (LISEEDLNGAALEGGDQ**N**ATGGHHHHHH) digested from sfGFP-DQNAT (predicted *m/z* 1090.5). Fragmentation with an isolation window that included the entire glycopeptide isotopic envelope with 30 eV revealed glycan fragment ions as well as intact peptide with fragmented glycan characteristic of the *C. jejuni* glycan. Highest intensity peaks are labeled and are +1 charge states unless otherwise indicated. **(B)** A triply-charged precursor ion (denoted with a blue diamond) was identified as the glycopeptide (NVGGDLDWPAAA**S**APQPGKPPR) digested from sfGFP-MOOR (predicted *m/z* 1202.9). Fragmentation with an isolation window that included the entire glycopeptide isotopic envelope with 30 eV also revealed characteristic glycan fragment ions and intact peptide with fragmented glycan characteristic of the *C. jejuni* glycan.

Highest intensity peaks are labeled and are +1 charge states unless otherwise indicated. Previous reports and glycosylation site amino acid mutation studies shown in **Supplementary Figure 10** strongly suggest that the glycan modification is on the bolded **N** and **S** residues within the sequons on sfGFP-DQNAT and sfGFP-MOOR glycopeptides, respectively.




**Supplementary Figure 12.** Characterization of *O*-linked glycosylation in CFGpS extracts. **(A)** α-His Western blot of PglO CFGpS reactions run with 20-minute CFPS times and **(B)** α-glycan blot corresponding to data in **(A)**. $g_0$ denotes aglycosylated acceptor protein and $g_1$ indicates glycoprotein. **(C)** Percent of acceptor protein converted to glycoprotein. Error bars represent standard deviation of three independent CFGpS reactions.




**Supplementary Figure 13**. Characterization of *N*-linked glycosylation with various acceptor proteins in CFGpS extracts using a $^{14}$C-L incorporation assay. CFGpS reactions were run in triplicate S12 and S30 extracts with a 20 min CFPS time **(A)** Autoradiogram of AcrA CFGpS reactions in triplicate S12 and S30 extracts and negative controls. $g_0$ denotes aglycosylated acceptor protein, $g_1$ indicates singly glycosylated glycoprotein and $g_2$ indicates doubly glycosylated glycoprotein. **(B)** Autoradiogram of MBP CFGpS reactions in triplicate S12 and S30 extracts and negative controls. $g_0$ denotes aglycosylated acceptor protein and $g_1$ indicates glycoprotein. **(C)** Autoradiogram of PD CFGpS reactions in triplicate S12 and S30 extracts and negative controls. $g_0$ denotes aglycosylated acceptor protein and $g_1$ indicates glycoprotein. **(D)**

Total soluble acceptor protein produced in triplicate CFGpS reactions run with S12 and S30 extracts as measured by scintillation counting of $^{14}C$-labeled proteins. Error bars represent standard deviation of three independent reactions. **(E)** Percent of acceptor protein glycosylated in CFGpS reactions run with S12 and S30 extracts for each acceptor protein tested as determined by densitometry of full-length product on autoradiograms. Error bars represent standard deviation of three independent reactions. AcrA glycoprotein was considered as the sum of singly and doubly glycosylated protein.

**Supplementary Table 1.** Information on proteins selected for extract enrichment in this study.

| Protein | Taxonomical origin | Predicted # TM helices | Function(s) | Size (kDa) | UniProt ID |
|---|---|---|---|---|---|
| PglB | *Campylobacter jejuni* | 13 | Catalyzes *N*-linked glycosylation | 82 | Q5HTX9 |
| PglO | *Neisseria gonorrhoeae* | 11 | Catalyzes *O*-linked glycosylation | 68 | Q5FA54 |
| NarX | *Escherichia coli* | 2 | Signal transduction for nitrate biosensing | 68 | P0AFA2 |
| Proteorhodopsin (PR) | Uncultured marine gamma proteobacterium EBAC3108 | 7 | Green light absorbing proteorhodopsin | 28 | Q9F7P4 |
| Cannabinoid receptor 1 (CB1) | *Homo sapiens* | 7 | G protein coupled receptor, molecular sensing | 54 | P21554 |
| STT3D | *Leishmania major* | 11 | Catalyzes *N*-linked glycosylation | 95 | E9AET9 |

**Supplementary Table 2**. Acceptor protein coding sequences used in CFGpS reactions.

| Sequon | Coding Sequence |
|---|---|
| pJL1-sfGFP-DQNAT | ATGAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGCGAT<br>GTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGATGCCACGATTGGCAAACT<br>GACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTGACCA<br>CCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCACGATTTCTT<br>TAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATGGCAAA<br>TATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAA<br>GGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCC<br>ATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGCAGTTTACCGTTCGCCA<br>TAACGTTGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGA<br>TGGTCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCC<br>GAACGAAAAAGGCACGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTAC<br>GCTAGGTGCGGCCGCAGAACAAAAACTCATCTCAGAAGAGGATCTGAATGGGGCCGCACTCG<br>AGGGTGGCGATCAGAACGCGACCGGCGGTCATCACCATCATCACCATTAA |
| pJL1-sfGFP AQNAT | ATGAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGCGAT<br>GTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGATGCCACGATTGGCAAACT<br>GACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTGACCA<br>CCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCACGATTTCTT<br>TAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATGGCAAA<br>TATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAA<br>GGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCC<br>ATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGCAGTTTACCGTTCGCCA<br>TAACGTTGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGA<br>TGGTCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCC<br>GAACGAAAAAGGCACGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTAC<br>GCTAGGTGCGGCCGCAGAACAAAAACTCATCTCAGAAGAGGATCTGAATGGGGCCGCACTCG<br>AGGGTGGCGCTCAGAACGCGACCGGCGGTCATCACCATCATCACCATTAA |
| pJL1-sfGFP-MOOR | ATGAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGCGAT<br>GTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGATGCCACGATTGGCAAACT<br>GACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTGACCA<br>CCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCACGATTTCTT<br>TAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATGGCAAA<br>TATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAA<br>GGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCC<br>ATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCAT<br>AACGTTGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGAT<br>GGTCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCG<br>AACGAAAAAGGCACGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTACG<br>GGCTCTTCTGGAGGGTCTGGCGATCCACGCAATGTGGGTGGGGATTTGGACTGGCCGGCGG<br>CAGCGAGTGCACCTCAACCCGGTAAACCTCCTCGTCATCACCACCATCATCACTAA |
| pJL1-sfGFP-MOORmut | ATGAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGCGAT<br>GTGAACGGTCACAAATTCAGCGTGCGTGGTGAAGGTGAAGGCGATGCCACGATTGGCAAACT<br>GACGCTGAAATTTATCTGCACCACCGGCAAACTGCCGGTGCCGTGGCCGACGCTGGTGACCA<br>CCCTGACCTATGGCGTTCAGTGTTTTAGTCGCTATCCGGATCACATGAAACGTCACGATTTCTT<br>TAAATCTGCAATGCCGGAAGGCTATGTGCAGGAACGTACGATTAGCTTTAAAGATGATGGCAAA<br>TATAAAACGCGCGCCGTTGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAA<br>GGCACGGATTTTAAAGAAGATGGCAATATCCTGGGCCATAAACTGGAATACAACTTTAATAGCC<br>ATAATGTTTATATTACGGCGGATAAACAGAAAAATGGCATCAAAGCGAATTTTACCGTTCGCCAT<br>AACGTTGAAGATGGCAGTGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGAT<br>GGTCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACGCAGACCGTTCTGTCTAAAGATCCG<br>AACGAAAAAGGCACGCGGGACCACATGGTTCTGCACGAATATGTGAATGCGGCAGGTATTACG<br>GGCTCTTCTGGAGGGTCTGGCGATCCACGCAATGTGGGTGGGGATTTGGACTGGCCGGCGG<br>CAGCGGGTGCACCTCAACCCGGTAAACCTCCTCGTCATCACCACCATCATCACTAA |

| | |
|---|---|
| pJL1-AcrA | ATGAAAGAAGAAGCACCAAAAATACAAATGCCGCCTCAACCTGTAACAACCATGAGTGCTAAAT<br>CTGAAGATTTACCACTTAGTTTTACTTACCCTGCTAAACTTGTCAGTGATTATGATGTCATTATAA<br>AACCTCAAGTTAGCGGCGTAATAGTAAATAAACTTTTTAAAGCTGGAGATAAGGTAAAAAAAGG<br>ACAAACATTATTTATTATAGAACAAGATAAATTTAAAGCTAGTGTTGATTCAGCTTACGGACAGG<br>CTTTAATGGCTAAGGCAACTTTCGAAAATGCAAGCAAGGATTTTAATCGTTCTAAAGCTCTTTTT<br>AGCAAAAGTGCAATCTCTCAAAAAGAATACGACTCTTCTCTTGCTACATTTAACAATTCAAAAGC<br>TAGTCTAGCAAGTGCTAGAGCACAGCTTGCAAATGCAAGAATTGATCTAGATCATACCGAGATA<br>AAAGCTCCTTTTGATGGTACTATAGGAGATGCTTTAGTTAATATAGGAGATTATGTAAGTGCTTC<br>AACAACTGAACTAGTTAGAGTTACAGATTTAAATCCTATTTACGCAGATTTCTTTATTTCAGATAC<br>AGATAAACTAAATTTAGTCCGCAATACTCAAAGTGGAAAATGGGATTTAGACAGCATTCATGCA<br>AATTTAAATCTTAATGGAGAAACCGTTCAAGGCAAACTTTATTTTATTGATTCGGTTATAGATGCT<br>AATAGTGGAACAGTAAAAGCCAAAGCCGTATTTGATAACAATAACTCAACACTTTTACCGGGTG<br>CTTTTGCAACAATTACCTCAGAAGGTTTTATACAAAAAAATGGCTTTAAAGTGCCTCAAATAGGT<br>GTTAAACAAGATCAAAATGATGTTTATGTTCTTCTCGTTAAAAATGGAAAAGTAGAAAAATCTTCT<br>GTACATATAAGCTACCAAAACAATGAATACGCCATTATTGACAAAGGATTGCAAAATGGCGATA<br>AAATCATTTTAGATAACTCTAAAAAAATTCAAGTTGGTAGCGAAGTTAAAGAAATTGGAGCACAA<br>CACCATCACCATCACCATTAA |
| pJL1-MBP-DQNAT | ATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCG<br>CTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATA<br>AACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGG<br>CACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAG<br>CGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTG<br>CTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCC<br>AAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGC<br>TGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATG<br>CGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGA<br>AAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA<br>CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGC<br>ATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGG<br>TCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAA<br>AGAGCTGGCGAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAA<br>TAAAGACAAACCGCTGGGTACCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCC<br>ACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGAT<br>GTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTG<br>TCGATGAAGCCCTGAAAGACGCGCAGACTCGTATCACCAAGCTAGGTGCGGCCGCAGAACAA<br>AAACTCATCTCAGAAGAGGATCTGAATGGGGCCGCACTCGAGGGTGGCGATCAGAACGCGAC<br>CGGCGGTCATCACCATCATCACCATTAA |
| pJL1-PD-DQNAT | ATGAAGTCAGACAAGATCATTATCGCGCATCGCGGTGCCTCCGGCTACTTGCCGGAGCATACT<br>TTGGAGAGCAAAGCCTTAGCATTTGCCCAACAAGCGGATTACTTGGAACAAGACTTGGCCATG<br>ACTAAGGATGGGCGTCTGGTGGTGATTCACGACCATTTTCTTGATGGTCTTACGGATGTAGCTA<br>AGAAGTTTCCGCATCGCCATCGTAAGGACGGGCGCTATTATGTAATCGATTTTACCTTGAAGGA<br>AATTCAAAGCTTGGAGATGACCGAAAACTTTGAGACGAAGGATGGGAAGCAAGCGCAAGTCTA<br>CCCGAATCGTTTTCCCTTATGGAAAAGCCATTTTCGCATCCATACTTTTGAGGATGAGATTGAG<br>TTTATCCAAGGGCTGGAAAAGTCTACTGGCAAGAAAGTTGGGATCTATCCTGAGATCAAGGCA<br>CCGTGGTTTCATCATCAAAACGGTAAGGACATCGCCGCTGAGACCTTAAAAGTACTGAAGAAG<br>TATGGCTACGATAAAAAGACTGACATGGTTTACCTTCAAACGTTCGATTTTAATGAGTTGAAACG<br>TATTAAAACAGAATTATTGCCCCAGATGGGGATGGATTTGAAGTTGGTTCAGTTAATCGCCTAC<br>ACCGACTGGAAGGAAACGCAGGAGAAGGATCCCAAAGGATACTGGGTGAATTATAATTATGAT<br>TGGATGTTCAAGCCAGGAGCAATGGCTGAAGTAGTAAAATATGCGGATGGTGTAGGCCCCGG<br>CTGGTATATGCTTGTGAACAAAGAAGAAAGCAAGCCTGATAATATCGTCTATACTCCCTTAGTC<br>AAAGAACTGGCCCAATATAACGTAGAAGTTCACCCTTACACCGTCCGTAAGGACGCCTTACCG<br>GAGTTTTTTACTGACGTCAATCAAATGTATGATGCCTTGCTGAACAAGAGTGGAGCGACAGGC<br>GTCTTCACCGACTTTCCAGACACTCTAGGTGCGGCCGCAGAACAAAAACTCATCTCAGAAGAG<br>GATCTGAATGGGGCCGCACTCGAGGGTGGCGATCAGAACGCGACCGGCGGTCATCACCATCA<br>TCACCATTAA |

**Supplementary Table 3.** Strains and plasmids used in this study.

| Strain/Plasmid | Description | Antibiotic Resistance | Reference |
|---|---|---|---|
| *E. coli* CLM24 | W3110, ΔWaaL | N/A | (1) |
| pJL1-sfGFP | pJL1 plasmid encoding superfolder GFP | Kan50 | (2) |
| pSF-CjPglB- | pSN18 derivative encoding *C. jejuni* PglB with C-terminal FLAG epitope tag | Carb100 | (3) |
| pSF-NgPglO | pSN18 derivative encoding *N. gonorrhoeae* PglO with C-terminal FLAG epitope tag | Carb100 | This study |
| pSF-EcNarX | pSN18 derivative encoding *E. coli* NarX with C-terminal FLAG epitope tag | Carb100 | This study |
| pSF-PR | pSN18 derivative encoding Uncultured marine gamma proteobacterium EBAC3108 proteorhodopsin with C-terminal FLAG epitope tag | Carb100 | This study |
| pSF-HsCB1 | pSN18 derivative encoding *H. sapiens* Cannabinoid receptor 1 with C-terminal FLAG epitope tag | Carb100 | This study |
| pSF-LmSTT3D | pSN18 derivative encoding *L. major* STT3D with C-terminal FLAG epitope tag | Carb100 | This study |
| pSF-sfGFP | pSN18 derivative encoding superfolder GFP with C-terminal FLAG epitope tag | Carb100 | This study |
| pMW07-pglΔB | pMW07 plasmid encoding *C. jejuni* protein glycosylation locus (pgl) with complete in-frame deletion of CjPglB | Cm34 | (3) |
| pJL1-sfGFP-DQNAT | pJL1 plasmid encoding superfolder GFP modified at the C-terminus with 30 amino acids containing an optimal DQNAT sequon followed by a 6x-His tag | Kan50 | This study, see **Table S2** for sequence |
| pJL1-sfGFP-AQNAT | pJL1 plasmid encoding superfolder GFP modified at the C-terminus with 30 amino acids containing a non-permissible AQNAT sequon followed by a 6x-His tag | Kan50 | This study, see **Table S2** for sequence |
| pJL1-sfGFP-MOOR | pJL1 plasmid encoding superfolder GFP modified at the C-terminus with 32 amino acids containing the minimum optimal *O*-linked recognition site (MOOR)[4] followed by a 6x-His tag | Kan50 | This study, see **Table S2** for sequence |
| pJL1-sfGFP-MOORmut | pJL1 plasmid encoding superfolder GFP modified at the C-terminus with 32 amino acids containing a non-permissible minimum optimal *O*-linked recognition site (MOORmut) followed by a 6x-His tag | Kan50 | This study, see **Table S2** for sequence |
| pJL1-AcrA | pJL1 plasmid encoding wt AcrA from *C. jejuni* containing 2 native internal glycosylation sites recognized by PglB and a C-terminal 6x-His tag | Kan50 | (5) |
| pJL1-MBP-DQNAT | pJL1 plasmid encoding *E. coli* maltose binding protein (MBP) modified at the C-terminus with 30 amino acids containing an optimal DQNAT sequon followed by a 6x-His tag | Kan 50 | This study, see **Table S2** for sequence. Note coding sequence of acceptor protein is used in a previous study with a different sequon (6). |
| pJL1-PD-DQNAT | pJL1 plasmid encoding *H. influenzae p*rotein D (PD) modified at the C-terminus with 30 amino acids containing an optimal DQNAT sequon followed by a 6x-His tag | Kan50 | This study, see **Table S2** for sequence. Note coding sequence of acceptor protein is used in a previous study with a different sequon (6). |

## Supplementary Methods

### *Zeta Potential Analysis*

Zeta potential measurements were performed in triplicate for 15 scans per measurement on a Zetasizer Nano ZS (Malvern Instruments Ltd.) using standard settings at room temperature and in disposable zeta potential cuvettes (Malvern Instruments Ltd., UK DTS1070).

## Supplementary References

[12–4]


1. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0500044102
2. Stark, J. C. *et al.* BioBits™ Bright: A fluorescent synthetic biology education kit. *Sci. Adv.* **4**, (2018).
3. Ollis, A. A., Zhang, S., Fisher, A. C. & DeLisa, M. P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* **10**, 816–822 (2014).
4. Pan, C. *et al.* Biosynthesis of conjugate vaccines using an O-linked glycosylation system. *MBio* **7**, (2016).
5. Schoborg, J. A. *et al.* A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnology and Bioengineering* (2017). doi:10.1002/bit.26502
6. Stark, J. C. *et al.* On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *bioRxiv* (2019). doi:10.1101/681841