In situ detection of protein interactions for recombinant therapeutic enzymes

Mojtaba Samoudi*[,1,2], Chih-Chung Kuo*[,2,3], Caressa M. Robinson*[,2,3], Km Shams-Ud-Doha[4], Song-Min Schinn[1,2], Stefan Kol[5], Linus Weiss[6], Sara Petersen Bjorn[5], Bjorn G. Voldborg[5], Alexandre Rosa Campos[4], Nathan E. Lewis[1,2,3,§]

[1] Dept of Pediatrics, University of California, San Diego

[2] Novo Nordisk Foundation Center for Biosustainability at UC San Diego

[3] Dept of Bioengineering, University of California, San Diego

[4] Sanford Burnham Prebys Medical Discovery Institute

[5] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark

[6] Dept of Biochemistry, Eberhard Karls University of Tübingen, Germany

[§] Corresponding author: Nathan E. Lewis

9500 Gilman Dr. MC 0760, Building BRF2, Room 4a16, La Jolla, CA 92093

Email: nlewisres@ucsd.edu, Tel: (858) 997 – 5844

Short title: In situ detection of protein interactions.

Abstract

Despite their therapeutic potential, many protein drugs remain inaccessible to patients since they are difficult to secrete. Each recombinant protein has unique physicochemical properties and requires different machinery for proper folding, assembly, and post-translational modifications (PTMs). Here we aimed to identify the machinery supporting recombinant protein secretion by measuring the protein-protein interaction (PPI) networks of four different recombinant proteins (SERPINA1, SERPINC1, SERPING1 and SeAP) with various PTMs and structural motifs using the proximity-dependent biotin identification (BioID) method. We identified PPIs associated with specific features of the secreted proteins using a Bayesian statistical model, and found proteins involved in protein folding, disulfide bond formation and N-glycosylation were positively correlated with the corresponding features of the four model proteins. Among others, oxidative folding enzymes showed the strongest association with disulfide bond formation, supporting their critical roles in proper folding and maintaining the ER stability. Knockdown of disulfide-isomerase PDIA4, a measured interactor with significance for SERPINC1 but not SERPINA1, led to the decreased secretion of SERPINC1, which relies on its extensive disulfide bonds, compared to SERPINA1, which has no disulfide bonds. Proximity-dependent labeling successfully identified the transient interactions supporting synthesis of secreted recombinant proteins and refined our understanding of key molecular mechanisms of the secretory pathway during recombinant protein production.

Keywords: Therapeutic proteins, Secretory pathway, BioID, Cell engineering, Disulfide bond.

Introduction

Therapeutic proteins are increasingly important for treating diverse diseases, including cancers, autoimmunity/inflammation, infectious diseases, and genetic disorders. For example, the plasma protein therapeutics market is expected to grow by $36 billion (USD) by 2024. Mammalian cells are the dominant production system due to their ability to perform PTMs that are required for drug safety and function (Jenkins, Murphy, & Tyther, 2008; Matasci, Hacker, Baldi, & Wurm, 2008). However, the complexities associated with the mammalian secretory machinery remains a bottleneck in recombinant protein production (Gutierrez et al., 2020). The secretory pathway machinery includes >575 gene products tasked with the synthesis, folding, PTMs, quality control, and trafficking of secreted proteins (SecPs) (Feizi, Gatto, Uhlen, & Nielsen, 2017; Lund et al., 2017; Novick, Ferro, & Schekman, 1981; Reynaud & Simpson, 2002). Numerous components of the secretory pathway (SecMs) have been engineered to increase the capacity of the secretion (Xiao, Shiloach, & Betenbaugh, 2014). However, the precision and efficiency of the mammalian secretory pathway results from the coordinated effort of these secretory machinery components including chaperones, modifying enzymes (e.g., protein disulfide isomerases and glycosyltransferases), and transporters within the secretory pathway. Overexpression of heterologous proteins in this tightly regulated and complex system could impact its functionality and homeostasis, resulting in adaptive responses that can impair both protein quantity and quality (Hussain, Maldonado-Agurto, & Dickson, 2014; Young, Yuraszeck, & Robinson, 2011). More importantly, variability in the structures and modifications of recombinant proteins could necessitate a customized secretion machinery to handle this diversity, but the secretory machinery of recombinant protein producing cells has not been adapted to facilitate the high titer secretion desired for most recombinant proteins. A previous study also showed human protein secretory pathway genes are expressed in a tissue-specific pattern to support the diversity of secreted proteins and their modifications (Feizi et al., 2017), suggesting that expression of several SecMs is regulated to support client SecPs in the secretory pathway. Unfortunately, the SecMs needed to support any given secreted protein remain unknown. Thus, there is a need to elucidate the SecMs that support the expression of different recombinant proteins with specific features. This can guide

mammalian synthetic biology efforts to engineer enhanced cells capable of expressing proteins of different kinds in a client-specific manner.

PPI networks are invaluable tools for deciphering the molecular basis of biological processes. New proximity dependent labeling methods such as BioID (Kim et al., 2014; Roux, Kim, Raida, & Burke, 2012) and APEX (Rhee et al., 2013) can identify weak and transient interactions in living cells, along with stable interactions. Furthermore, BioID offers a high-throughput approach for systematic detection of intracellular PPIs occurring in various cellular compartments and has been used to characterize PPI networks and subcellular organization (Varnaitė & MacNeill, 2016). BioID relies on expressing a protein of interest fused to a promiscuous biotin ligase (BirA) that can biotinylate the proximal interactors in nanometer-scale labeling radius (Kim et al., 2014). For example, this approach has mapped protein interactions at human centrosomes and cilia (Firat-Karalar & Stearns, 2015; Gupta et al., 2015), focal adhesions (Dong et al., 2016), nuclear pore (Kim et al., 2014) and ER membrane-bound ribosomes (Hoffman, Chen, Zheng, & Nicchitta, 2019) . BirA has also been used in proximity-specific ribosome profiling to reveal principles of ER co-translational translocation (Jan, Williams, & Weissman, 2014). Here we used BioID2, an improved smaller biotin ligase for BioID (Kim et al., 2016; Varnaitė & MacNeill, 2016), to explore how the SecMs involved vary for different secreted therapeutic proteins (Fig. 1). Specifically, BioID2 was employed to identify SecMs that interact with three SERPIN-family proteins (SERPINA1: treatment for Alpha-1-antitrypsin deficiency, SERPINC1: treatment for Hereditary antithrombin deficiency, and SERPING1: treatment for acute attacks of hereditary angioedema) and secreted embryonic alkaline phosphatase (SeAP), which is a truncated form of Alkaline Phosphatase, Placental Type (ALPP). These proteins vary in their PTMs (e.g., glycosylation, disulfide bond and residue modifications) and have different amino acid sequences that consequently form different local motifs. Using a Bayesian statistical model, we identified the critical PPIs that are positively correlated with each protein feature. Identification of these PPIs will refine our understanding of how the secretory pathway functions during the expression of the recombinant proteins and introduce novel targets for secretory pathway engineering in a client specific manner.

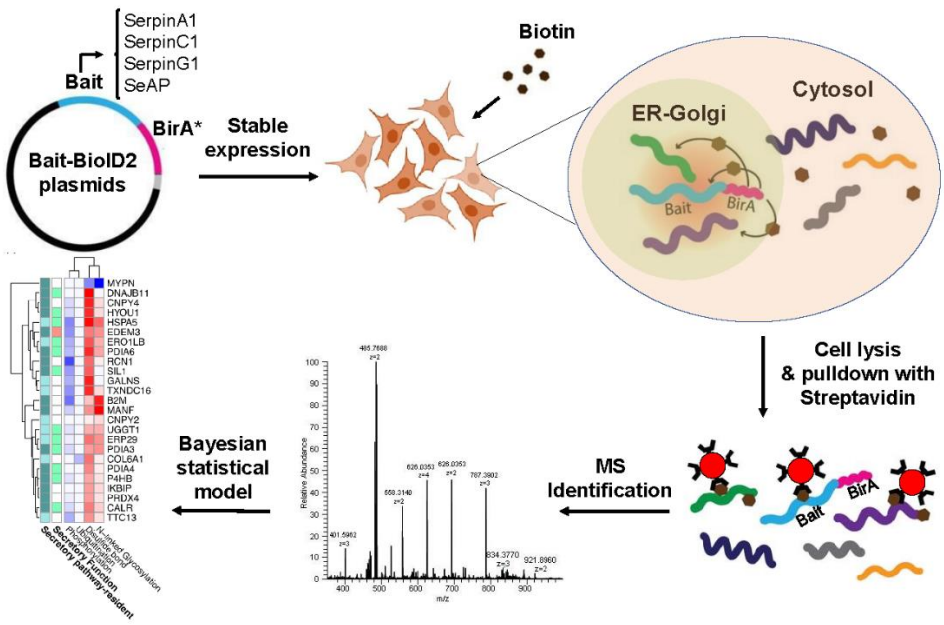

Figure 1. Flowchart of the BioID2 application to detect in situ interactions supporting therapeutic proteins secretion.

Materials and Methods

Molecular cloning and generation of stable cell lines

All plasmids used in this study were constructed by PCR and Gibson isothermal assembly. The expression ORFs, hereafter named bait-BirA, were constructed by fusing BioID2 to the C-terminal of each model protein (with a glycine-serine linker added between) and a 3XFLAG tag at C-terminal to simplify the immuno detection. ORFs were inserted into pcDNA5/FRT (Invitrogen), which allows targeted integration of the transgenes into the host genome. Gibson assembly primers were designed by SnapGene software and used to amplify the corresponding fragments and vectors with long overlapping overhangs, which were then assembled using Gibson Assembly Master Mix (NEB). To obtain secretable BioID2 (without any bait protein), Gibson assembly was employed to fuse the signal peptide of SERPINC1 gene to the N-terminal of BirA (hereafter referred to as Signal-BirA). Assembly products were transformed to the chemically competent *E. coli*, and recombinant plasmids were verified by restriction digestion and sequencing. For all experiments, Flp-In 293 cells (Invitrogen) were cultured in DMEM media supplemented with fetal bovine serum (10 %) and antibiotics (penicillin, 100 U mL−1 and streptomycin, 100 μg mL−1) and maintained at 37 °C under 5 % CO2. For generating stable cell lines, Flp-In 293 cells

were seeded in 6 well plates at a density of 0.5×10^6 cells per well the day before transfection. Cells were then co-transfected with each pcDNA5/FRT vector containing expression cassette and pOG44 plasmid using Lipofectamine® 2000 according to the manufacturer's directions. After recovery from transfection, cells were grown in DMEM containing 10% FBS, 1% PenStrep, and 150 µg/mL Hygromycin B to select hygromycin-resistant cells. Individual resistant colonies were isolated, pooled, and seeded in 24-well plates for further scaling up and screened for expression of the fusion proteins by Western Blotting.

Immunofluorescence

Recombinant HEK293 cells expressing BioID2 fusions were grown in complete medium supplemented with 50 µM biotin on coverslips until 70% confluent. Cells were then fixed in PBS containing 4% PFA for 10 min at room temperature. Blocking was performed by incubating fixed cells with 1% BSA and 5% normal goat serum in PBST. Anti-flag mouse monoclonal antibody-Dylight 650 conjugate (Thermofisher), targeting the bait-BirA, and streptavidin-DyLight 594 conjugate (Thermofisher), targeting the biotinylated proteins, were diluted at 1:300 and 1:1000 in blocking buffer, respectively and incubated with fixed cells for 30 minutes at room temperature. Cells were then washed, counterstained with DAPI, mounted on the slide using antifade vectashield mountant, and imaged using Leica SP8 Confocal with Lightning Deconvolution. Colocalization quantification was performed for the deconvolved images using Fiji's (ImageJ 1.52p) Coloc_2 analysis tool between the 650 (anti-flag) and 594 (Streptavidin) channels (Schindelin, Rueden, Hiner, & Eliceiri, 2015). This tool generates a comprehensive report for evaluating pixel intensity colocalization of two channels by various methods such as Pearson's Coefficient (range: -1.0 to 1.0), Manders' Colocalization Coefficients (MCC, range: 0 to 1.0), and Li's Intensity Correlation Quotient (IQC, range: -0.5 to 0.5) (Li, 2004; Manders, Verbeek, & Aten, 1993). Background pixel intensity was subtracted using Fiji's rolling ball algorithm and a region of interest (ROI). Thresholds were determined using Coloc_2's bisection method, which is further used to adjust for background. Above threshold metrics were reported.

RNAi knockdown experiment

An orthogonal siRNA approach called esiRNA (Sigma Aldrich) was employed to target PDIA4, PDIA6, and ERp44. HEK293 cell expressing SERPINC1-BirA and SERPINA1-BirA were seeded at $1X10^5$

cells/well in 12-well plates with complete medium and reverse transfected with 144 ng of the appropriate PDI specific esiRNA, Luciferase esiRNA as a negative control, or KIF11 esiRNA as positive control using Lipofectamine RNAiMAX (Invitrogen). All transfections were performed according to the manufacturer's guidelines. Each experiment was done in triplicates and targeted gene knockdown by esiRNA was allowed to occur for 96 hrs. Culture supernatants and cell pellets were then harvested, clarified by low-speed centrifugation, and then aliquoted and stored at −80°C for further experiments.

Western blotting

To validate the secretion of bait-BirA proteins, supernatants of cultures expressing fusion proteins were collected, centrifuged to remove cell debris, and 30 µl were loaded on SDS-PAGE gel for electrophoresis. The resolved proteins were then transblotted to nitrocellulose membranes using the Trans-Blot Turbo Transfer System from Bio-RAD. The membrane was blocked with 5% skim milk in TBST and probed with HRP-conjugated anti-flag mouse monoclonal antibody (Thermofisher) diluted at 1:10000 in the blocking buffer. The membrane was washed, and Clarity Western ECL Substrate was added. Proteins' bonds were visualized using G:Box Gel Image Analysis Systems (SYNGENE). For staining of intracellular biotinylated proteins, cells were grown in complete medium supplemented with 50 µM biotin, lysed by RIPA buffer, and protein content was quantified using Bradford assay. 20 ug of total protein was loaded, resolved and transblotted as described earlier. The membrane was blocked by 3% BSA in TBST and probed with HRP-conjugated streptavidin diluted in blocking buffer at 1:2000 ratio. For visualizing the proteins' bands, the same Clarity Western ECL Substrate was used.

Quantitative western blots were used to determine knockdown (KD) efficiency as well as impact on SERPIN secretion. For KD efficiency, cell pellets from each KD experiment were lysed with RIPA buffer, and approximately 25µg of protein lysate was loaded onto the SDS-PAGE gel for PDIA4, PDIA6, and ERp44 experiments. Resolved proteins were transblotted onto separate nitrocellulose membranes as described earlier. Each membrane was blocked with Intercept (TBS) Blocking Buffer from LI-COR and probed with Rabbit anti PDIA4 (1:2000), PDIA6 (1:2000), or ERp44 (1:2500) monoclonal antibodies, respectively. A housekeeping protein in each lysate was targeted for normalization with either a Mouse anti Alpha-tubulin (1:20,000) or anti Beta-actin (1:10,000) monoclonal antibodies. For visualization, IR

imaging was utilized by staining with LI-COR Goat anti Mouse conjugated to 680 (1:15,000) and anti-Rabbit conjugated to 800 (1:15,000). Images were taken and analyzed with Image Studio Lite Version 5.2 for relative quantification. Impact on SERPIN secretion was determined from the aliquots of clarified supernatants from esiRNA transfected cultures using the same blotting method as above. 30 µl of supernatants were loaded on SDS-PAGE gel, and transblotted onto nitrocellulose membrane. Transblotted membrane was probed using Rabbit anti Flag (Proteintech, 1:800) followed by Goat anti Rabbit 800 (1:15,000). KD efficiency and the effect of PDI knockdown on secretion of the model proteins was measured in comparison with the negative control of each cell line transfected with Luciferase esiRNA as negative control (see above).

Mass Spectrometry

Cells were grown in 245 mm plates (one plate per biological replicate in triplicate) to approximately 70% confluence in complete media and then incubated for 24 h with 50 µM biotin. Cells were harvested and washed twice in cold PBS, lysed with vigorous shaking (20 Hz) in 8M urea, 50mM ammonium bicarbonate lysis buffer, extracted proteins were centrifuged at 14,000 x g to remove cellular debris and quantified by BCA assay (Thermo Scientific) as per manufacturer recommendations. Affinity purification of biotinylated proteins was carried out in a Bravo AssayMap platform (Agilent) using AssayMap streptavidin cartridges (Agilent), and the bound proteins were subjected to on-cartridge digestion with mass spec grade Trypsin/Lys-C Rapid digestion enzyme (Promega, Madison, WI) at 70ºC for 2h. Digested peptides were then desalted in the Bravo platform using AssayMap C18 cartridges and the organic solvent was removed in a SpeedVac concentrator prior to LC-MS/MS analysis. Dried peptides were reconstituted with 2% acetonitrile, 0.1% formic acid, and analyzed by LC-MS/MS using a Proxeon EASY nanoLC system (Thermo Fisher Scientific) coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Peptides were separated using an analytical C18 Acclaim PepMap column 0.075 x 500 mm, 2µm particles (Thermo Scientific) in a 93-min linear gradient of 2-28% solvent B (80% acetonitrile, 0.1% formic acid) at a flow rate of 300nL/min. The mass spectrometer was operated in positive data-dependent acquisition mode. MS1 spectra were measured with a resolution of 70,000, an AGC target of 1e6 and a mass range from 350 to 1700 m/z. Up to 12 MS2 spectra per duty cycle were triggered, fragmented by

HCD, and acquired with a resolution of 17,500 and an AGC target of 5e4, an isolation window of 1.6 m/z and a normalized collision energy of 25. Dynamic exclusion was enabled with a duration of 20 sec.

MS data Analysis

All mass spectra were analyzed with MaxQuant software (Tyanova, Temu, & Cox, 2016) version 1.5.5.1. MS/MS spectra were searched against the Homo sapiens Uniprot protein sequence database (version January 2018) and GPM cRAP sequences (commonly known protein contaminants). Precursor mass tolerance was set to 20ppm and 4.5ppm for the first search where initial mass recalibration was completed and for the main search, respectively. Product ions were searched with a mass tolerance 0.5 Da. The maximum precursor ion charge state used for searching was 7. Carbamidomethylation of cysteines was searched as a fixed modification, while oxidation of methionines and acetylation of protein N-terminal were searched as variable modifications. Enzyme was set to trypsin in a specific mode and a maximum of two missed cleavages was allowed for searching. The target-decoy-based false discovery rate (FDR) filter for spectrum and protein identification was set to 1%. Enrichment of proteins in streptavidin affinity purifications were calculated as the ratio of intensity. To remove the systematic biases introduced during various steps of sample processing and data generation, dataset were normalized using the LOESS method (Smyth, n.d.) integrated into Normalyzer (Chawade, Alexandersson, & Levander, 2014). Perseus software (Tyanova & Cox, 2018) was employed for data preparation, filtering, and computation of differential protein abundance. The DEP package (Zhang et al., 2018) was used to explore whether missing values in the dataset are biased to lower intense proteins. Left-censored imputation was performed using random draws from shifted distribution. A Student's t-test with a multi-sample permutation-based correction for an FDR of 0.05 was employed to identify differentially expressed proteins using log2 transformed data.

Detection of significant interactions

The threshold for significant interactions was determined using the known secretory pathway components as a gold standard. We set the cutoffs for FDR at 0.1 and removed all interactors with negative fold changes, as this optimizes the enrichment of known secretory pathway components among the significant interactors. The enrichment for two independent secretory pathway-related gene sets also

peaked around the cutoffs set through the gene set of known secretory pathway components, suggesting the optimal cutoffs are robust to the gold standards chosen.

Estimation of preferential interaction between protein features and interactors with a Bayesian modeling framework

To identify patterns of interactions between individual bait-BirA proteins and their interactors, we first obtained and summarized several protein features across model proteins. The protein properties considered include shared structural motifs (Blatch & Lässle, 1999), known sites of PTM from Uniprot and phosphosite (Blatch & Lässle, 1999; Hornbeck et al., 2015). The complete protein feature composition for each of the model proteins is illustrated in (Fig. S4). Given the interactions between the SecPs and their interactors, we can predict the important structural features implicated in the interactions between the SecPs and a given SecM. The interactions between the BirA-fused samples and the secretory pathway interactors can be pooled according to shared properties of the SecPs to reveal interdependencies between components of the secretory pathway and their products.

To test if some secretory machinery components preferentially interact with certain protein features, we first calculated the effective total frequency ($\delta_{f,g}$) of interactions between each feature-gene pair (f,g) by going through every SecP in our data and counting the number of times this feature occurs in a bait-BirA protein p (f_p).

$$\delta_{f,g} = \sum_{p \in \text{secPs}} \mathbf{f}_p \cdot \mathbb{I}_{\textbf{interact}}(p,g) \qquad \text{(bait-BirA protein: } p\text{; interactor: } g\text{)}$$

$$\mathbb{I}_{\textbf{interact}}(p,g) := \begin{cases} 1 & \text{if p and g interact} \\ 0 & \text{otherwise.} \end{cases}$$

The number f_p is added to the total frequency, $\delta_{f,g}$, only when p and g interact (when $\mathbb{I}_{interact}(p,g) = 1$). This effective interaction frequency, $\delta_{f,g}$ essentially linked the features pooled across the model proteins to a given interactor g, taking us closer to estimating interaction affinity between a feature and an interactor. To further account for SecP and feature promiscuity, we implement an estimate of the tendency for f and g to interact. The interaction synergy $s_{f,g}$, the tendency for an interactor g to interact with a feature f more than expected by chance, was estimated by a Bayesian modeling approach. More specifically, we seek

to decouple the interaction synergy from the observed interaction frequency $\delta_{f,g}$ by modeling the interaction frequency $\delta_{f,g}$ with a Poisson regression.

$$\delta_{f,g} \sim \text{Poisson}(\lambda_{f,g}) \qquad \text{(Interaction frequency between } f, g \text{ follows a Poisson distribution with mean } \lambda_{f,g})$$
$$\log \lambda_{f,g} = \alpha + \mu_f + \mu_g + s_{fg}$$

Interaction effects:

$$\alpha \sim \text{Normal}(0, 1) \quad \text{(Average interaction frequency across all } f, g \text{ pairs)}$$
$$\mu_f \sim \text{Normal}(0, \sigma_f) \quad \text{(Feature promiscuity for } f\text{)}$$
$$\mu_g \sim \text{Normal}(0, \sigma_g) \quad \text{(Interactor promiscuity for } g\text{)}$$
$$s_{fg} \sim \text{Normal}(0, \sigma_s) \quad \text{(Interaction synergy between } f, g\text{)}$$

Hyper priors:

$$\sigma_f \sim \text{Exponential}(1)$$
$$\sigma_g \sim \text{Exponential}(1)$$
$$\sigma_s \sim \text{Exponential}(1)$$

The mean for the Poisson distribution is parameterized by the sum of feature promiscuity (number of SecMs *g* connected to a feature) μ_f , interactor promiscuity (number of SecPs with a feature *f* interacting with a SecM) μ_g , the interaction synergy $s_{f,g}$ and an intercept variable. The feature promiscuity μ_f quantifies the probability of a feature *f* to partake in an interaction with any secMs. Likewise, the interactor promiscuity μ_g measures the tendency for a secM *g* to interact with any features. After subtracting out the μ_f and μ_g from the log-transformed mean $\lambda_{f,g}$ for the Poisson distribution, we arrive at the coefficient of interest-- the interaction synergy $s_{f,g}$ between *f* and *g*. It quantifies the degree to which *f* and *g* interact more than by random chance. In previous work, this approach has correctly estimated epistasis intensity(Shen et al., 2017). To better regularize the promiscuities, their Bayesian priors are all normally distributed around 0, with their variances parameterized by the hyper priors σ_f, σ_g and σ_s which follow an exponential distribution. The intercept α is parameterized by a standard normal distribution. We used the *rethinking* R package (McElreath, 2020) to construct the model and sample the coefficients.

Results

BioID can successfully tag proteins colocalized with secreted proteins

We first investigated if intracellular PPIs between each SecP and their supporting SecMs can be measured using the BioID method. To do this, each bait-BirA was expressed in HEK293 cells using the Flp-In™ system (see materials and methods) for targeted integration of the transgenes into the same genomic locus to ensure comparable transcription rates of each transgene. Variations in mRNA level caused by random integration can trigger adaptive response such as the unfolded protein response in

some cell lines which reciprocally alters the active PPIs network involved in the secretion. We observed successful secretion of bait-BirA proteins into culture supernatant, evaluated by Western blot (Fig. 2a). Thus, the BirA fusion did not prevent secretion of the model proteins, and it is expected that they enter the secretory pathway where they are processed and packaged for secretion. We also verified the biotinylation profile by western blot for each cell line in the presence and absence of biotin. The biotinylation profile of the bait-BirA cells is different when biotin is added to the culture with substantial increased biotinylation of specific proteins, while no obvious change is observed for WT (Fig. 2b), suggesting that BioID2 successfully tagged specific proteins within the cells.

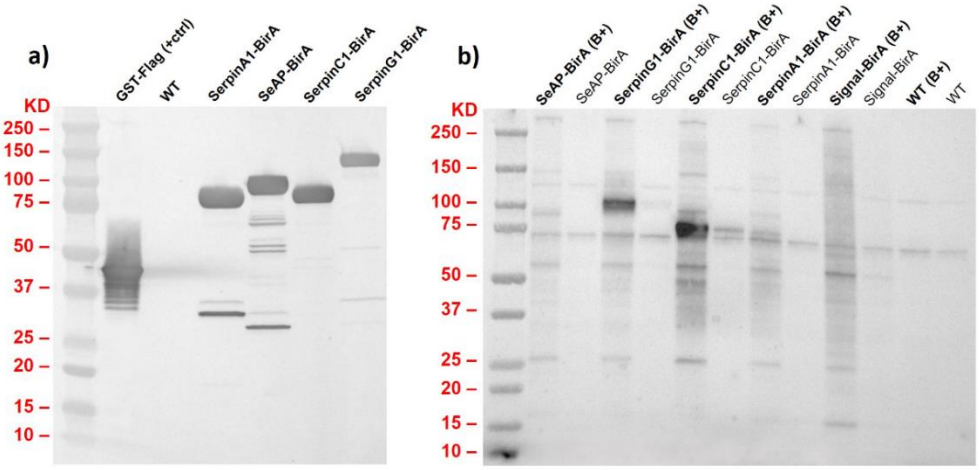

Figure 2. Expression of bait-BirA proteins results in a substantial increase in biotinylated proteins. a) Successful secretion of the bait-BirA proteins into the culture supernatant was evaluated by Western blot using HRP-anti-flag antibody. b) The immunoblotting biotinylation profiling of the model proteins and WT control in HEK293 cells with HRP-streptavidin. When the BirA domain was fused to the model proteins, biotin addition led to the biotinylation of a subset of proteins (B+) which are not seen in WT or absence of biotin. This demonstrates that the BioID labeling system tags interactions as secreted proteins are synthesized and trafficked through the secretory pathway. A few endogenously biotinylated proteins appear in the absence of biotin and in the WT.

Colocalization of the bait-BirA proteins and the biotinylated proteins was then studied by multicolor co-immunofluorescence microscopy to test whether biotinylated proteins are actual partners of the model proteins. The results demonstrated successful labeling of the interactors by BirA through colocalization of the biotin-labeled proteins and bait-BirA, while WT did not show increased biotinylation under the same experimental conditions (Fig. 3). To quantify the colocalizations, we calculated different colocalization

metrics (see methods) from the images and compared to the WT (Table 1 and Fig. S1), and the results confirmed the specificity of the BioID labeling system to tag the proximal proteins (Li's ICQ value closer to 0.5 and Pearson's R value and Manders' Colocalization Coefficients closer to 1 demonstrate a dependent protein staining pattern between the red and green channels).

Table 1. Colocalization metrics between the 650 (anti-flag) and 594 (Streptavidin) channels in Fig. S1. There is little correlation reported for the WT, whereas there is much higher correlation for clones containing BirA.

Cells	Pearson's R value	Li's ICQ value	Manders' Coeff. (tM_g,tM_r)
WT	0.22	0.219	0.812, 0.518
SERPINA1-BirA	0.74	0.318	0.962, 0.888
SERPING1-BirA	0.73	0.260	0.999, 0.914
SERPINC1-BirA	0.88	0.392	0.991, 0.982
SeAP-BirA	0.93	0.392	0.952, 0.935
Signal-BirA	0.71	0.308	0.841, 0.791

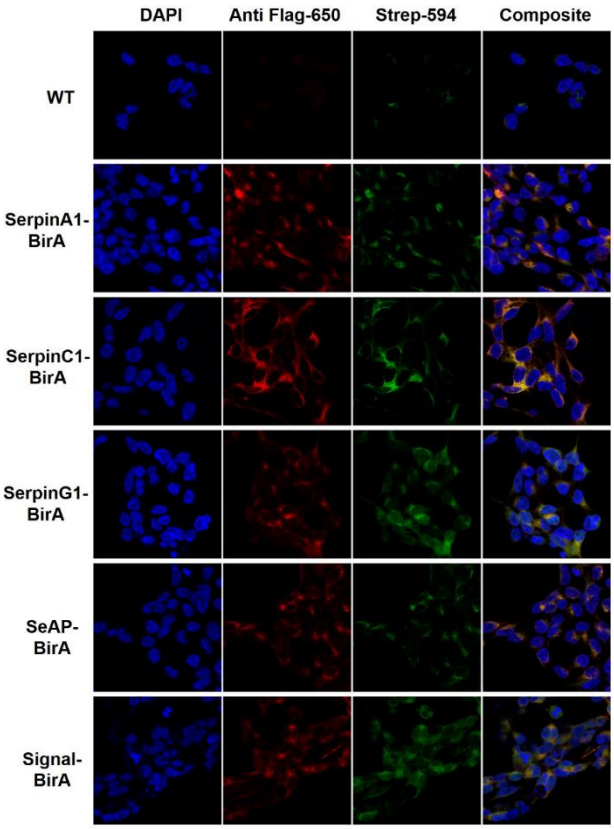

Figure 3. Bait-BirA fusion proteins are colocalized with biotin-staining. Co-Immunofluorescence demonstrated the intracellular colocalization of the biotin-labeled proteins (stained with Streptavidin-Dylight 594 and illustrated in green color) and bait-BirA (stained with anti-flag monoclonal antibody-Dylight 650 and illustrated in red color), while WT did not show increased biotinylation under the same experimental conditions.

WT cells revealed endogenous biotinylation landscape

After successful tagging of the proximal proteins we aimed to identify the interactions with each bait protein. For this, cells were lysed, and biotinylated proteins were purified using streptavidin. Purified proteins were digested by trypsin, and peptides were subjected to LC-MS/MS, and the biotinylated proteins in samples were mapped using MaxQuant (Tyanova et al., 2016). Differentially biotinylated proteins were then identified in each sample compared to the WT using Perseus (Tyanova & Cox, 2018). When implemented in a WT control cell line, we identified proteins that are biotinylated endogenously along with bona fide interactors. These include proteins that bind biotin as a cofactor such as carboxylases which is problematic with streptavidin-based protein detection (Tytgat et al., 2015). Although the extent to which the general endogenous biotinylation has not been systematically quantified, the biotinylated proteins isolated from the WT sample showed considerable overlap with interacting proteins detected in other model protein samples, suggesting endogenous biotinylation may be more pervasive than previously believed. Among the top differentially biotinylated proteins in bait-BirA samples, the bait proteins showed the highest log-fold change (LFC) (Fig. 4 and Table S1). This observation is expected because the bait protein is a potential substrate for BirA located in the closest vicinity of the enzyme and is considered as evidence to show the biotinylation system is working properly in the cell.

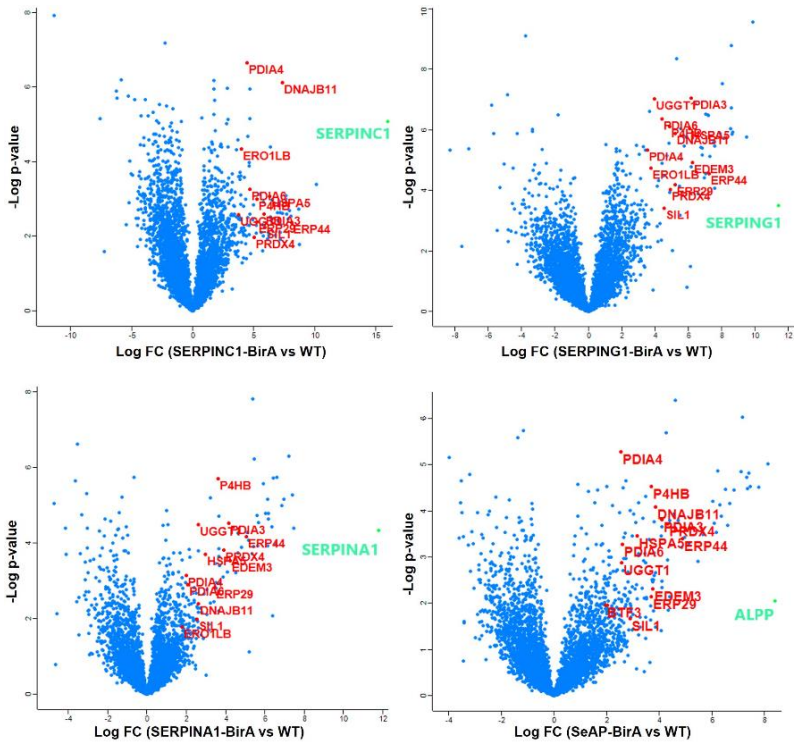

Figure 4. Dozens of proteins show significantly increased biotinylation after expression of bait-BirA proteins. Volcano plot showing the distribution of the quantified biotinylated proteins by MS according to p-value and fold change. As depicted the bait-protein significantly showed the highest fold change compared to WT almost in all cases, indicating the capability of the BioID labeling system to tag the in vivo interactions within the live cells. The key interactors involved in disulfide bond and N-glycosylation formation (see text) are highlighted in red. SeAP is a truncated form of Alkaline phosphatase, placental type (ALPP).

Interactors are enriched for secretory pathway components and co-secreted proteins

In theory, the model proteins would come in frequent contact with members of the secretory pathway, and other co-secreted proteins. To determine if our setup captures the secretory pathway-related proteins more than by random chance, we analyzed the enrichment for 3 independent secretory pathway-relevant gene sets at various fold change and p-value thresholds. We saw a significant enrichment for the secretory pathway machinery, secretory-resident and co-secreted proteins among probable PPIs across all model protein samples (Fig. 5a and Fig. S3), with peak enrichment occurring at the top 100-300 most significant interactors by positive fold change. This corresponds to a significance cutoff of a fold change of 3 or greater enrichment and an adjusted p-value < 0.1, in model proteins compared to WT control (Fig. S2 for all thresholds and Fig. 5b for all significant interactions) for each secreted protein (see methods). The secretory machinery components are more enriched among the top 300 hits for all model proteins than other co-secreted proteins, suggesting more frequent interactions between the secretory pathway machinery and their products than the crosstalk between co-secreted proteins. Probable PPIs detected in all model proteins (n=19) and hits shared among all SERPIN gene products are significantly enriched for proteins involved in protein folding (Fig. 5b). Indeed, molecular chaperones are highly promiscuous when assisting protein folding due to their inherent flexibility (Mayer, 2010). Apart from the shared interactions, PPIs for each model protein differ substantially. Thus, the question remained if these private interactions correspond to unique properties of each model protein.

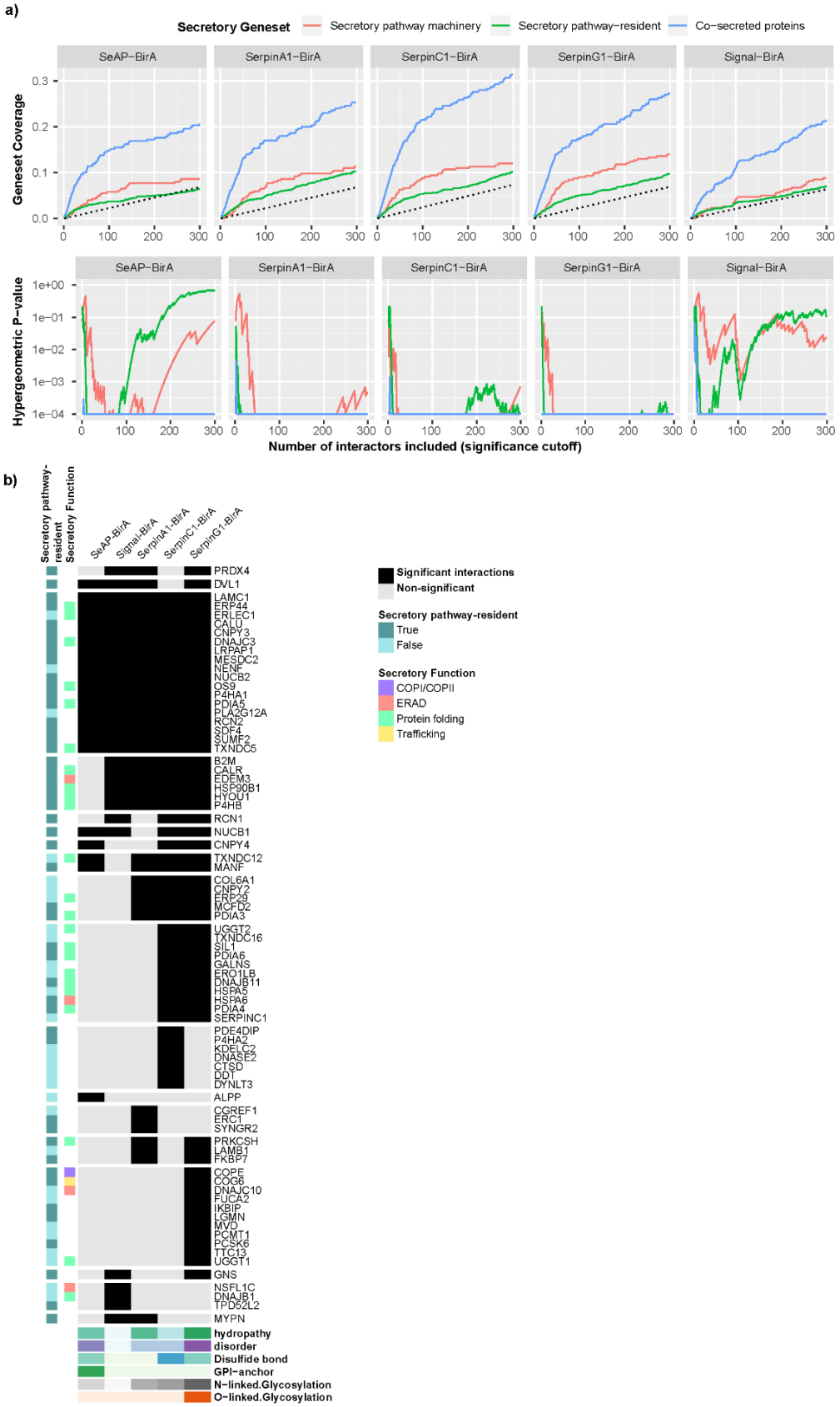

Figure 5. Interacting proteins are enriched for secretory pathway machinery. (a) To determine if significant interactions enrich for secretory pathway-related genes, we performed an iterative enrichment analysis in which we

included the interactors with the greatest fold changes first and iteratively added interactors with lower fold changes. The y-axis indicates the overall coverage of 3 secretory pathway-related gene sets and the x-axis the significance cutoffs (rank ordered by fold change). The coverage of the gene set (top) along with their corresponding hypergeometric enrichment p-value (bottom) are shown, denoting the probability of obtaining an overlap larger than the one observed if no enrichment existed. The top 300 hits for each secretable BirA sample (Fig. S3 for all hits) showed significant enrichment of the secretory pathway components and co-secreted proteins among the most significant hits for all samples except Signal-BirA (which is a lone secreted BirA and not a mammalian secreted protein). (b) Quantified interactions between interactors (y-axis) and the model proteins (x-axis), where the shadowed entries indicate significant interactions. The features of the model proteins, detailed in Fig. S4, are summarized for each model protein on the bottom panel and the secretory pathway attributes for the interactors are labeled on the left.

Private interactors reflect post-translational and structural features of model proteins

PPIs in the secretory pathway mediate the folding, modification and transportation of secreted proteins (Bonifacino & Glick, 2004; Calakos, Bennett, Peterson, & Scheller, 1994; Ikawa et al., 1997; Pearl & Prodromou, 2006; Watanabe et al., 2019). Incidentally, co-expression analysis has linked certain PTMs across the secretome to the expression of their responsible enzymes. For example, PDIs are consistently upregulated in tissues secreting disulfide-rich proteins (Feizi et al., 2017). As the bait-BirA proteins differ in structural composition and PTMs (Fig. S4), we wondered if bait-BirA proteins with shared features have higher affinity for specific interactors. More specifically, we hypothesize that proteins requiring a specific PTM would preferentially interact with the secretory machinery components responsible for the PTM synthesis. To test if such preferential interaction exists, we summarized various PTM and structural properties across model proteins and analyzed their associations with the corresponding secretory machinery using a Bayesian modeling framework (see methods). Among the studied PTMs, bait-BirA proteins with disulfide bonds and N-linked glycans demonstrated higher affinity towards specific interactions (Fig. 6a) that are known to help secretion of proteins with the corresponding PTMs. Thus, we analyzed the detected interactions associated with glycosylation, disulfide bond addition, and protein folding.

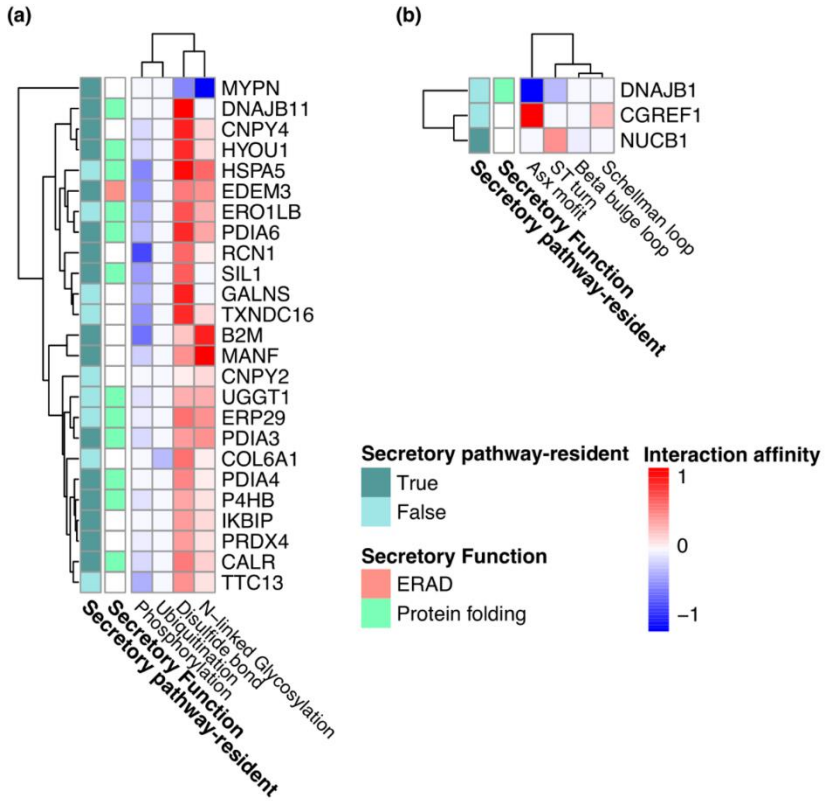

Figure 6. Detected interactors correlate with protein features. Interactors were associated with specific (a) PTMs and (b) structural features of model proteins. The heatmap shows the standardized interaction affinities estimated between certain interactors and PTMs or structural features across all model proteins (see methods). Only interactors having significant associations with model protein features are shown.

Proteins with increased glycosylation are associated with quality control pathways

We detected significant interactions in the Calnexin/Calreticulin cycle and related processes for more heavily glycosylated proteins (Fig. 6a). For example, the glycosylated baits interacted with calreticulin (CALR), a calcium-binding chaperone that promotes folding, oligomeric assembly, and quality control of glycoproteins in the ER(Nauseef, McCormick, & Clark, 1995). They also interacted with UGGT1, which recognizes glycoproteins with minor folding defects and reglucosylates single N-glycans near the misfolded part of the protein. Reglucosylated proteins are then recognized by CALR for recycling to the ER and refolding or degradation (Ferris, Jaber, Molinari, Arvan, & Kaufman, 2013). Two members of the PDI family, PDIA3 and ERp29, which form a complex with calreticulin/calnexin, also showed association with N-glycosylated baits (Fig. 6a) suggesting their role in glycoprotein folding and quality control. Calnexin/Calreticulin-PDIA3 complexes promote the oxidative folding of nascent polypeptides (Sakono,

Seko, Takeda, & Ito, 2014) and ERp29 promotes isomerization of peptidyl-prolyl bonds to attain the native polypeptide structure (Sakono et al., 2014; Tannous, Pisoni, Hebert, & Molinari, 2015). Proteins with chaperone activity, such as HSPA5 (Fig. 6a), were also found to interact with N-linked glycan-containing bait-BirA. HSPA5 is a component of the glycoprotein quality-control (GQC) system which recognizes glycoproteins with amino acid substitutions, and targets them for ER-associated degradation (ERAD) (Ferris, Kodali, & Kaufman, 2014). EDEM3, another interactor associated with the N-glycan containing proteins (Fig. 6a), is a glycosyltransferase involved in ERAD mediated degradation of glycoproteins by catalyzing mannose trimming from Man8GlcNAc2 to Man7GlcNAc2 in N-glycans (Ninagawa et al., 2014). Given that most of these molecular chaperones and enzymes are involved in ERAD-mediated degradation of the misfolded glycoproteins, these findings suggest the quality control pathways are critical for synthesizing and secreting proteins with N-linked glycans.

Disulfide bond formation is rate-limiting in protein secretion

Several members of the PDI family including P4HB, PDIA3, PDIA4 and PDIA6 significantly interacted with model-BirAs containing more disulfide bonds (Fig. 6a). These enzymes catalyze the formation, breakage and rearrangement of disulfide bonds through the thioredoxin-like domains (Kozlov, Määttänen, Thomas, & Gehring, 2010). The identification of various PDIs highlights the importance of the oxidative folding enzymes in protein folding and maintaining stability that can limit the efficiency of protein secretion. The proteins with more disulfide bonds also interact with major ER chaperones HSPA5 and DNAJB11, a co-chaperone of HSPA5, that play a key role in protein folding and quality control in the ER lumen (Ng, Watowich, & Lamb, 1992; Yu, Haslam, & Haslam, 2000), highlighting their important role in secretion of the disulfide bond enriched proteins. The PDI, ERp44 showed the highest association (LFC > 8, Fig. S5) with disulfide bond-enriched proteins i.e. SERPINC1 and SERPING1 although it also showed considerable interaction with SERPINA1-BirA and a less strong interaction with SeAP-BirA. ERp44 mediates the ER retention of the oxidoreductase Ero1α (an oxidoreductin that reoxidizes P4HB to enable additional rounds of disulfide formation) through the formation of reversible mixed disulfides (Anelli et al., 2003). Hence, the association of ERp44 and disulfide bonds highlights the importance of the thiol-mediated ER protein retention in disulfide bond formation, particularly when the secretory

pathway is loaded with proteins with more disulfide bonds. In addition, ERO1LB, PRDX4 and SIL1 were ER-localized enzymes that were associated with disulfide bond formation. ERO1LB efficiently reoxidizes P4HB (Mezghrani et al., 2001), PRDX4 couples hydrogen peroxide catabolism with oxidative protein folding by reducing hydrogen peroxide (Zito, 2013), and SIL1 can reverse HSPA5 cysteine oxidation which alters its chaperone activity to cope with suboptimal folding conditions (Siegenthaler, Pareja, Wang, & Sevier, 2017). The identification of these oxidoreductase enzymes highlights the importance of ER redox homeostasis in disulfide bond formation and protecting cells from the consequences of misfolded proteins.

To validate the importance of specific PDI interactions that showed highest fold change and specificity of effect on productivity of proteins with more disulfide bonds, we knocked down PDIA4, PDIA6, or ERp44 in cells expressing either SERPINC1-BirA or SERPINA1-BirA, using an orthogonal RNAi approach, i.e. esiRNAs (Kittler, Heninger, Franke, Habermann, & Buchholz, 2005). HEK293 cells expressing SERPINC1-BirA or SERPINA1-BirA were transfected with esiRNA against PDIA4, PDIA6, ERp44, or LUC and KIF11 as negative/positive controls, respectively. Knockdown experiments successfully resulted in >70% reduction in cellular expression of the targeted PDIs (Fig. 7a). We observed a 42% reduction in SERPINC1-BirA secretion following knockdown of PDIA4 (p-value = 0.032), meanwhile a significantly lesser reduction of SERPINA1-BirA secretion was seen (Fig. 7b). If differences in KD efficacy between experiments are accounted for, normalization of SERPINs secretion by KD efficiencies resulted in 54% reduction of SERPINC1-BirA secretion following knockdown of PDIA4 (p-value = 0.026) while no significant reduction was seen in SERPINA1-BirA secretion (Fig. 7c). Neither knockdown of PDIA6 nor ERp44 produced a significant reduction in either SERPINC1-BirA nor SERPINA1-BirA secretion (Fig. 7b & 7c). BioID analysis indicated that PDIA4 interacts significantly with SERPINC1-BirA but not with SERPINA1-BirA, while PDIA6 did not show considerable interaction with SERPINA1 nor SERPINC1 (Fig. S5). Therefore, PDIA4 may work as a private interactor for SERPINC1 secretion as recognized by knockdown results. PDIA6 could potentially work as a private interactor for SERPING1 secretion as predicted by BioID (Fig. S5). ERp44 had a higher quantified fold change interaction with SERPINC1-BirA,

but it also showed significant interaction with SERPINA1-BirA, SERPING1-BirA and less strongly with SeAP-BirA, suggesting ERp44 may function in a more public manner.

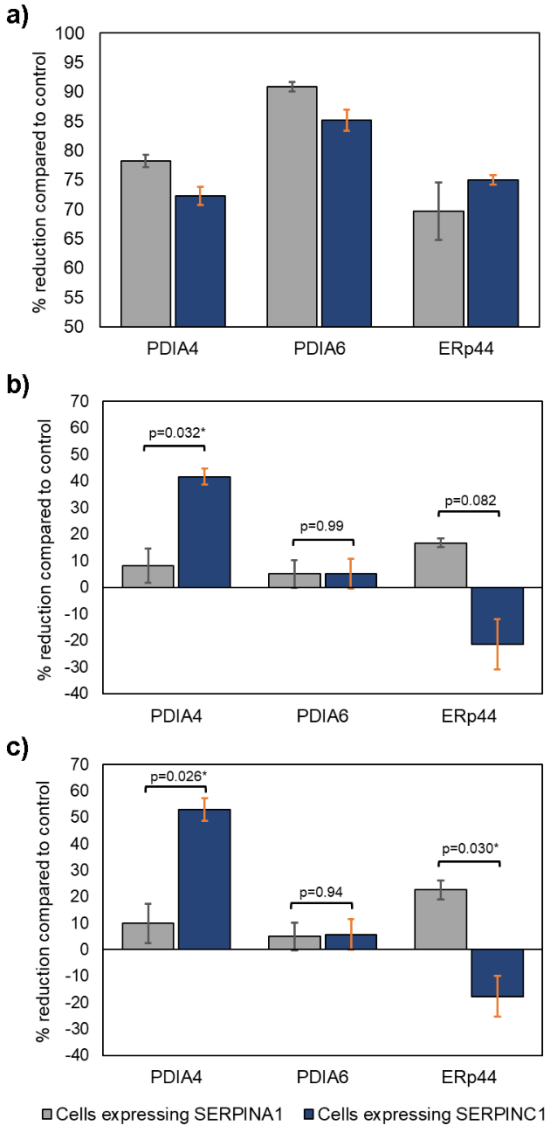

Figure 7. Effects of esiRNA mediated knockdown of isomerases PDIA4, PDIA6, and ERp44 on SERPIN secretion. Quantitative western blots were used to find the relative abundance of each PDI and SERPIN after esiRNA transfection in both cells expressing SERPINA1 and SERPINC1. We have reported these results as (a) KD efficiency of each PDI measured by percent reduction of PDI abundance compared to the negative control, (b) the effect of each PDI KD on SERPINA1 and SERPINC1 secretion, and (c) the normalized effects of PDI KD on SERPIN secretion using the KD efficiency, assuming greater KD results would result in a greater loss of SERPIN secretion. SERPINC1, which has a multitude of disulfide PTMs, shows a greater reduction in SERPIN production in PDIA4

KD experiments compared to SERPINA1, which does not have as many disulfide bridges in its structure. This is consistent with expected interactions found with BioID analysis.

Identified PPIs are associated with structural motifs on bait proteins

In addition to PTMs, the Bayesian modeling framework found associations between SecP structural features and the SecMs (Fig. 6b). For example, model proteins depleted in the asx motif (Golovin & Henrick, 2008) showed higher tendency to interact with DNAJB1, a cytosolic molecular chaperone of the HSP40 family, possibly through retrotranslocation of the misfolded model proteins in ERAD. The asx motif impacts N-glycan occupancy of Asn-X-Thr/Ser sites, depending on the ability of the peptide to adopt an Asx-turn motif (Imperiali, Shannon, & Rickert, 1992; Imperiali, Shannon, Unno, & Rickert, 1992). As another example, NUCB1, a chaperone-like amyloid binding protein that prevents protein aggregation (Bonito-Oliva, Barbash, Sakmar, & Graham, 2017), interacted more strongly with our proteins with more ST-turns (Fig. 6b). ST turns occur frequently at the N-termini of α-helices (Doig, Stapley, Macarthur, & Thornton, 2008) and are regarded as helix capping features which stabilize α-helices in proteins (Aurora & Rosee, 1998). Thus, the enriched interaction of the NUCB1 with St-turn suggests that it can help stabilize folding of protein with a predominant α-helical secondary structure.

Discussion

BioID has been used to profile the proteome of different cellular compartments and molecular complex systems (Varnaitė & MacNeill, 2016). However, this is the first time that BioID has been used to identify the proteome of the secretory pathway during recombinant protein expression. Numerous PPIs guiding the folding, modifications, and trafficking of the secreted and membrane proteins through the secretory pathway are transient (Nyfeler, Michnick, & Hauri, 2005; Schreiber, Haran, & Zhou, 2009), and therefore, cannot be captured by conventional methods such as co-immunoprecipitation. Consistent with previous studies (Sears, May, & Roux, 2019), these results showed the BioID can detect weak and transient interactions *in situ,* and therefore it is a powerful approach to study luminal processes involved in protein secretion. We found that disulfide bridge formation enzymes showed the strongest association with bait proteins enriched in disulfide bonds, supporting their critical roles in protein secretion and maintaining ER stability. A previous study on difficult to express (DTE) monoclonal antibodies showed less recognition

by PDI impairs disulfide bridge formation within the antibody light chain (LC) which can initiate the intracellular degradation by the ubiquitin proteasome system via ERAD (Mathias et al., 2020). Thus, insufficient interaction between the secreted proteins with enriched disulfide bonds and PDIs can limit secretion efficiency and serve as a rate-limiting step for protein production. In another study, the tissue specific analysis of SecMs expression showed a positive correlation between the expression of P4HB and PDIA4 and liver tissue where numerous disulfide bond enriched proteins are secreted by hepatocytes (Feizi et al., 2017). These observations are clear evidence that suggests the tissue-specific fine-tuning of the PDI family expression level in response to the enrichment of the disulfide sites. Together, these results showed PDIs are actively involved in adaptive responses and secretion of proteins with dominant disulfide bonds that are crucial for restoring ER stability, and therefore, yielding the recombinant proteins. Given the associations between the SecMs and the features of the model proteins, we also hypothesized that SecMs preferentially interacting with bait-BirA proteins that carry certain structural features may be essential for the secretion of those proteins. While evidence linking SecMs to the structural motifs is lacking, many molecular chaperones selectively interact with certain sequence and structural elements to favor the particular folding pathways (Gidalevitz, Stevens, & Argon, 2013). For example, chaperones of the HSP70 family evolved to bind extended β strand peptides; interestingly, the associations identified between chaperones and asx motif and ST turn represent a novel association for further study.

While we show BioID works well for studying the synthesis of secreted proteins, we acknowledge that biotin-based methods have some limitations as well. Biotin is actively imported into the cytoplasm of cells and can freely diffuse to the nucleus, but it may not be as accessible in the secretory pathway, thus reducing labeling efficacy in that compartment (Kim & Roux, 2016). Here, we showed this challenge is not an insurmountable issue, in that the BioID2 construct successfully identified many expected luminal interactions. BioID2 requires less biotin supplementation and exhibits enhanced labeling of proximate protein (Kim et al., 2016), allowing for BioID to be introduced to new systems where biotinylation supplementation may not be easily accomplished (Sears et al., 2019). More recently, two promiscuous mutants of biotin ligase, TurboID and miniTurbo, have been developed to catalyze proximity labeling even with much greater efficiency (Branon et al., 2020) and therefore, can be considered as an effective

method when proximal labeling of the endomembrane organelles is desired. It is also possible that the coverage of biotin labeling may be limited in the early processes in the secretory pathway due to the maturation process of the biotin ligase, i.e. biotinylation only can occur once the biotin ligase has been fully functional. Our results did reveal that there was coverage of the early secretory pathway interactions, but the possibility of missing interactions cannot be ruled out. So, BioID should be considered as a complementary method to other PPIs and systems biology approaches for complete characterization of endomembrane interactions.

In summary, we demonstrate here an approach to identify the protein interactions that synthesize and support secreted proteins, and thus define the product-specific secretory pathway. The identification of such machinery opens avenues for mammalian synthetic biology, wherein biotherapeutic production hosts can be rationally engineered to improve the titer and quality of diverse proteins in a client specific manner.

Acknowledgments

This work was supported by generous funding from the Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark (NNF10CC1016517, BGV) and from NIGMS (R35 GM119850, NEL). Light microscopy was performed at the UCSD School of Medicine Microscopy Core, which is supported by an NINDS P30 grant (NS047101). The authors would also like to thank Tune Wulff for support on pilot assays.

Conflict of Interest

The authors declare no competing interests.

References

Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., … Sitia, R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. *The EMBO Journal*, *22*(19), 5015–5022.

Aurora, R., & Rosee, G. D. (1998). Helix capping. *Protein Science*, Vol. 7, pp. 21–38. https://doi.org/10.1002/pro.5560070103

Blatch, G. L., & Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *21*(11), 932–939.

Bonifacino, J. S., & Glick, B. S. (2004). The Mechanisms of Vesicle Budding and Fusion. *Cell*, Vol. 116, pp. 153–166. https://doi.org/10.1016/s0092-8674(03)01079-1

Bonito-Oliva, A., Barbash, S., Sakmar, T. P., & Graham, W. V. (2017). Nucleobindin 1 binds to multiple types of pre-fibrillar amyloid and inhibits fibrillization. *Scientific Reports*, *7*, 42880.

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., … Ting, A. Y. (2020). Author Correction: Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology*, *38*(1), 108.

Calakos, N., Bennett, M. K., Peterson, K. E., & Scheller, R. H. (1994). Protein-protein interactions contributing to the specificity of intracellular vesicular trafficking. *Science*, *263*(5150), 1146–1149.

Chawade, A., Alexandersson, E., & Levander, F. (2014). Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *Journal of Proteome Research*, *13*(6), 3114–3120.

Doig, A. J., Stapley, B. J., Macarthur, M. W., & Thornton, J. M. (2008). Structures of N-termini of helices in proteins. *Protein Science*, Vol. 6, pp. 147–155. https://doi.org/10.1002/pro.5560060117

Dong, J.-M., Tay, F. P.-L., Swa, H. L.-F., Gunaratne, J., Leung, T., Burke, B., & Manser, E. (2016). Proximity biotinylation provides insight into the molecular composition of focal adhesions at the nanometer scale. *Science Signaling*, *9*(432), rs4.

Feizi, A., Gatto, F., Uhlen, M., & Nielsen, J. (2017). Human protein secretory pathway genes are expressed in a tissue-specific pattern to match processing demands of the secretome. *Npj Systems Biology and Applications*, Vol. 3. https://doi.org/10.1038/s41540-017-0021-4

Ferris, S. P., Jaber, N. S., Molinari, M., Arvan, P., & Kaufman, R. J. (2013). UDP-glucose:glycoprotein glucosyltransferase (UGGT1) promotes substrate solubility in the endoplasmic reticulum. *Molecular Biology of the Cell*, *24*(17), 2597–2608.

Ferris, S. P., Kodali, V. K., & Kaufman, R. J. (2014). Glycoprotein folding and quality-control mechanisms in protein-folding diseases. *Disease Models & Mechanisms*, *7*(3), 331–341.

Firat-Karalar, E. N., & Stearns, T. (2015). Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. *Methods in Cell Biology*, *129*, 153–170.

Gidalevitz, T., Stevens, F., & Argon, Y. (2013). Orchestration of secretory protein folding by ER chaperones. *Biochimica et Biophysica Acta*, *1833*(11), 2410–2424.

Golovin, A., & Henrick, K. (2008). MSDmotif: exploring protein sites and motifs. *BMC Bioinformatics*, *9*, 312.

Gupta, G. D., Coyaud, É., Gonçalves, J., Mojarad, B. A., Liu, Y., Wu, Q., … Pelletier, L. (2015). A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell*, *163*(6), 1484–1499.

Gutierrez, J. M., Feizi, A., Li, S., Kallehauge, T. B., Hefzi, H., Grav, L. M., … Lewis, N. E. (2020). Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications*, *11*(1), 68.

Hoffman, A. M., Chen, Q., Zheng, T., & Nicchitta, C. V. (2019). Heterogeneous translational landscape of the endoplasmic reticulum revealed by ribosome proximity labeling and transcriptome analysis. *The Journal of Biological Chemistry*, *294*(22), 8942–8958.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, *43*(Database issue), D512–D520.

Hussain, H., Maldonado-Agurto, R., & Dickson, A. J. (2014). The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. *Biotechnology Letters*, *36*(8), 1581–1593.

Ikawa, M., Wada, I., Kominami, K., Watanabe, D., Toshimori, K., Nishimune, Y., & Okabe, M. (1997). The putative chaperone calmegin is required for sperm fertility. *Nature*, *387*(6633), 607–611.

Imperiali, B., Shannon, K. L., & Rickert, K. W. (1992). Role of peptide conformation in asparagine-linked glycosylation. *Journal of the American Chemical Society*, Vol. 114, pp. 7942–7944. https://doi.org/10.1021/ja00046a068

Imperiali, B., Shannon, K. L., Unno, M., & Rickert, K. W. (1992). Mechanistic proposal for asparagine-linked glycosylation. *Journal of the American Chemical Society*, Vol. 114, pp. 7944–7945.

https://doi.org/10.1021/ja00046a069

Jan, C. H., Williams, C. C., & Weissman, J. S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, *346*(6210), 1257521.

Jenkins, N., Murphy, L., & Tyther, R. (2008). Post-translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals. *Molecular Biotechnology*, Vol. 39, pp. 113–118. https://doi.org/10.1007/s12033-008-9049-4

Kim, D. I., Birendra, K. C., Zhu, W., Motamedchaboki, K., Doye, V., & Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), E2453–E2461.

Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., & Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell*, *27*(8), 1188–1196.

Kim, D. I., & Roux, K. J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends in Cell Biology*, *26*(11), 804–817.

Kittler, R., Heninger, A.-K., Franke, K., Habermann, B., & Buchholz, F. (2005). Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nature Methods*, *2*(10), 779–784.

Kozlov, G., Määttänen, P., Thomas, D. Y., & Gehring, K. (2010). A structural overview of the PDI family of proteins. *The FEBS Journal*, *277*(19), 3924–3936.

Li, Q. (2004). A Syntaxin 1, G o, and N-Type Calcium Channel Complex at a Presynaptic Nerve Terminal: Analysis by Quantitative Immunocolocalization. *Journal of Neuroscience*, Vol. 24, pp. 4070–4081. https://doi.org/10.1523/jneurosci.0346-04.2004

Lund, A. M., Kaas, C. S., Brandl, J., Pedersen, L. E., Kildegaard, H. F., Kristensen, C., & Andersen, M. R. (2017). Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Systems Biology*, Vol. 11. https://doi.org/10.1186/s12918-017-0414-4

Manders, E. M. M., Verbeek, F. J., & Aten, J. A. (1993). Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy*, Vol. 169, pp. 375–382.

https://doi.org/10.1111/j.1365-2818.1993.tb03313.x

Matasci, M., Hacker, D. L., Baldi, L., & Wurm, F. M. (2008). Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discovery Today. Technologies*, *5*(2-3), e37–e42.

Mathias, S., Wippermann, A., Raab, N., Zeh, N., Handrick, R., Gorr, I., … Otte, K. (2020). Unraveling what makes a monoclonal antibody difficult-to-express: From intracellular accumulation to incomplete folding and degradation via ERAD. *Biotechnology and Bioengineering*, Vol. 117, pp. 5–16. https://doi.org/10.1002/bit.27196

Mayer, M. P. (2010). Gymnastics of molecular chaperones. *Molecular Cell*, *39*(3), 321–331.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. CRC Press.

Mezghrani, A., Fassio, A., Benham, A., Simmen, T., Braakman, I., & Sitia, R. (2001). Manipulation of oxidative protein folding and PDI redox state in mammalian cells. *The EMBO Journal*, *20*(22), 6288–6296.

Nauseef, W. M., McCormick, S. J., & Clark, R. A. (1995). Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase. *The Journal of Biological Chemistry*, *270*(9), 4741–4747.

Ng, D. T., Watowich, S. S., & Lamb, R. A. (1992). Analysis in vivo of GRP78-BiP/substrate interactions and their role in induction of the GRP78-BiP gene. *Molecular Biology of the Cell*, *3*(2), 143–155.

Ninagawa, S., Okada, T., Sumitomo, Y., Kamiya, Y., Kato, K., Horimoto, S., … Mori, K. (2014). EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step. *The Journal of Cell Biology*, *206*(3), 347–356.

Novick, P., Ferro, S., & Schekman, R. (1981). Order of events in the yeast secretory pathway. *Cell*, *25*(2), 461–469.

Nyfeler, B., Michnick, S. W., & Hauri, H.-P. (2005). Capturing protein interactions in the secretory pathway of living cells. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(18), 6350–6355.

Pearl, L. H., & Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone

machinery. *Annual Review of Biochemistry*, *75*, 271–294.

Reynaud, E. G., & Simpson, J. C. (2002). Navigating the secretory pathway: conference on exocytosis membrane structure and dynamics. *EMBO Reports*, *3*(9), 828–833.

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, *339*(6125), 1328–1331.

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology*, *196*(6), 801–810.

Sakono, M., Seko, A., Takeda, Y., & Ito, Y. (2014). PDI family protein ERp29 forms 1:1 complex with lectin chaperone calreticulin. *Biochemical and Biophysical Research Communications*, *452*(1), 27–31.

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, *82*(7-8), 518–529.

Schreiber, G., Haran, G., & Zhou, H.-X. (2009). Fundamental aspects of protein-protein association kinetics. *Chemical Reviews*, *109*(3), 839–860.

Sears, R. M., May, D. G., & Roux, K. J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods in Molecular Biology* , *2012*, 299–313.

Shen, J. P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., … Mali, P. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nature Methods*, *14*(6), 573–576.

Siegenthaler, K. D., Pareja, K. A., Wang, J., & Sevier, C. S. (2017). An unexpected role for the yeast nucleotide exchange factor Sil1 as a reductant acting on the molecular chaperone BiP. *eLife*, *6*. https://doi.org/10.7554/eLife.24141

Smyth, G. K. (n.d.). limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pp. 397–420. https://doi.org/10.1007/0-387-29362-0_23

Tannous, A., Pisoni, G. B., Hebert, D. N., & Molinari, M. (2015). N-linked sugar-regulated protein folding

and quality control in the ER. *Seminars in Cell & Developmental Biology*, *41*, 79–89.

Tyanova, S., & Cox, J. (2018). Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods in Molecular Biology* , *1711*, 133–148.

Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*(12), 2301–2319.

Tytgat, H. L. P., Schoofs, G., Driesen, M., Proost, P., Van Damme, E. J. M., Vanderleyden, J., & Lebeer, S. (2015). Endogenous biotin-binding proteins: an overlooked factor causing false positives in streptavidin-based protein detection. *Microbial Biotechnology*, *8*(1), 164–168.

Varnaitė, R., & MacNeill, S. A. (2016). Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics*, *16*(19), 2503–2518.

Watanabe, S., Amagai, Y., Sannino, S., Tempio, T., Anelli, T., Harayama, M., … Inaba, K. (2019). Zinc regulates ERp44-dependent protein quality control in the early secretory pathway. *Nature Communications*, *10*(1), 603.

Xiao, S., Shiloach, J., & Betenbaugh, M. J. (2014). Engineering cells to improve protein expression. *Current Opinion in Structural Biology*, Vol. 26, pp. 32–38. https://doi.org/10.1016/j.sbi.2014.03.005

Young, C. L., Yuraszeck, T., & Robinson, A. S. (2011). Decreased Secretion and Unfolded Protein Response Upregulation. *Methods in Enzymology*, pp. 235–260. https://doi.org/10.1016/b978-0-12-385928-0.00014-6

Yu, M., Haslam, R. H., & Haslam, D. B. (2000). HEDJ, an Hsp40 co-chaperone localized to the endoplasmic reticulum of human cells. *The Journal of Biological Chemistry*, *275*(32), 24984–24992.

Zhang, X., Smits, A. H., van Tilburg, G. B., Ovaa, H., Huber, W., & Vermeulen, M. (2018). Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature Protocols*, *13*(3), 530–550.

Zito, E. (2013). PRDX4, an endoplasmic reticulum-localized peroxiredoxin at the crossroads between enzymatic oxidative protein folding and nonenzymatic protein oxidation. *Antioxidants & Redox Signaling*, *18*(13), 1666–1674.