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

* Equal contribution
Correspondence to Nathan E. Lewis, nlewisres@ucsd.edu

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. Knock down of ERP44, a measured interactor with the highest fold change, led to the decreased secretion of SERPINC1, which relies on its extensive 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, Protein-Protein Interaction, 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 [1]. Mammalian cells are the dominant production system due to their ability to perform post-translational modifications (PTMs) that are required for drug safety and function [2,3]. However, the complexities associated with the mammalian secretory machinery remains a bottleneck in recombinant protein production [4].

The secretory pathway machinery includes >575 gene products tasked with the synthesis, folding, post-translational modification (PTM), quality control, and trafficking of secreted proteins (SecPs) [5–8]. The precision and efficiency of the mammalian secretory pathway results from the coordinated effort of these secretory machinery components (SecMs) 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 [9,10]. 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 [5], 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 synthetic biology efforts to engineer enhanced cells capable of expressing proteins of different kinds in a client-specific manner.

Protein-protein interaction (PPI) networks are invaluable tools for deciphering the molecular basis of biological processes. Traditionally, PPI discovery has leveraged diverse tools, such as yeast 2-hybrid or immunoprecipitation followed by mass spectrometry. These approaches are either conducted in a non-native state or require stronger, more stable interactions. However, new proximity dependent labeling methods such as BioID [11,12] and APEX [13] 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 [14,15]. BioID relies on expressing a protein of interest fused to a promiscuous biotin ligase (BirA) that releases activated biotin-AMP in the absence of substrate. Biotin-AMP covalently modifies the primary amines of proximal proteins within a nanometer-scale labeling radius

[12]. Biotinylated proximal proteins are identified by isolation with streptavidin, followed by tandem mass spectrometry (MS/MS). For example, this approach has mapped protein interactions at human centrosomes and cilia [16,17], focal adhesions [18], nuclear pore [12] and ER membrane-bound ribosomes [19] .

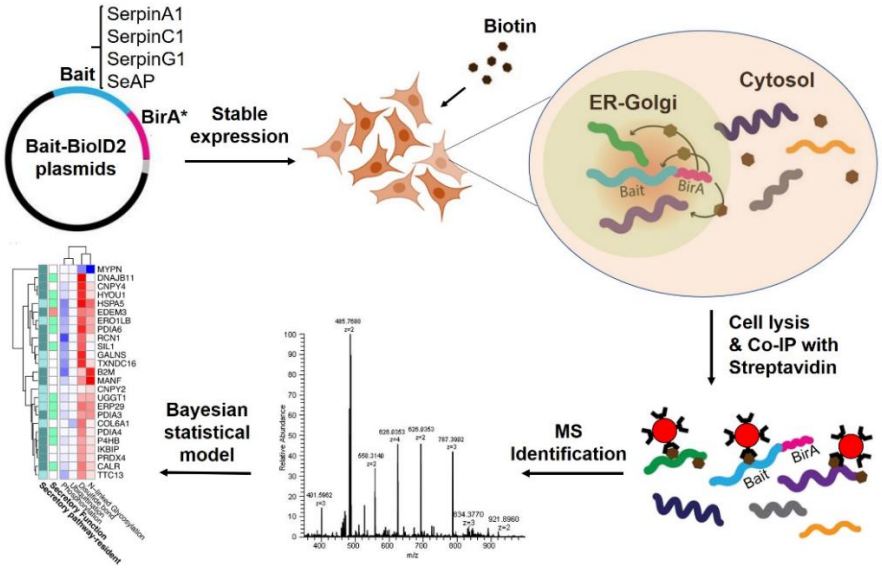


Figure 1. Flowchart of the BioID application to detect in situ interactions for therapeutic proteins secretion.

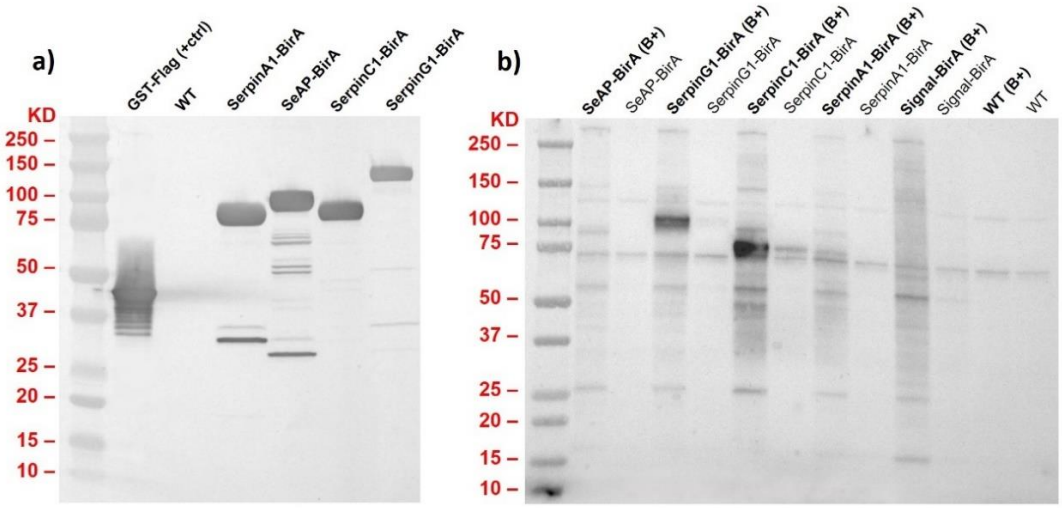
Here we used BioID2, an improved smaller biotin ligase for BioID [14,20], to explore how the SecMs involved vary for different secreted therapeutic proteins (Fig. 1). Specifically, BioID2 was employed to identify SecMs that interact with three SERPIN-family proteins (SERPINA1: treatment for Alpha-1-antitrypsin 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) gene. Interactions are measured while the proteins are trafficked through the ER and Golgi. 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. Thus, 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 bottlenecks in the secretory pathway during the expression of the recombinant proteins and introduce novel targets for secretory pathway engineering in a client specific manner.

Results

BioID can successfully tag proteins colocalized with secreted proteins

We first investigated if intracellular protein-protein interactions (PPIs) between each SecP and their supporting SecMs can be measured using the BioID method. To do this, each bait protein was fused to the BirA domain (bait-BirA) and expressed in HEK293 cells. The Flp-In™ system (see materials and methods) was chosen for targeted integration of the transgenes into the same genomic locus to ensure comparable transcription rates of each transgene from a Flp-In™ expression vector. Variations in mRNA level caused by random integration can trigger adaptive 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 change the intracellular localization of the model proteins, and it is expected that they enter the secretory pathway where they are processed and packaged for secretion. This is important since fusing large protein domains can potentially result in intracellular mislocalization of the proteins, which would lead to erroneous proximal protein labeling.



Fugure2. 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.

We verified the biotinylation profile by western blot for each cell line in the presence and absence of biotin. The biotinylation profile of the recombinant 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. Colocalization of the bait-BirA proteins and the biotinylated proteins was 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 biotinylated proteins and bait-BirA, while WT did not show increased biotinylation (Fig. 3).

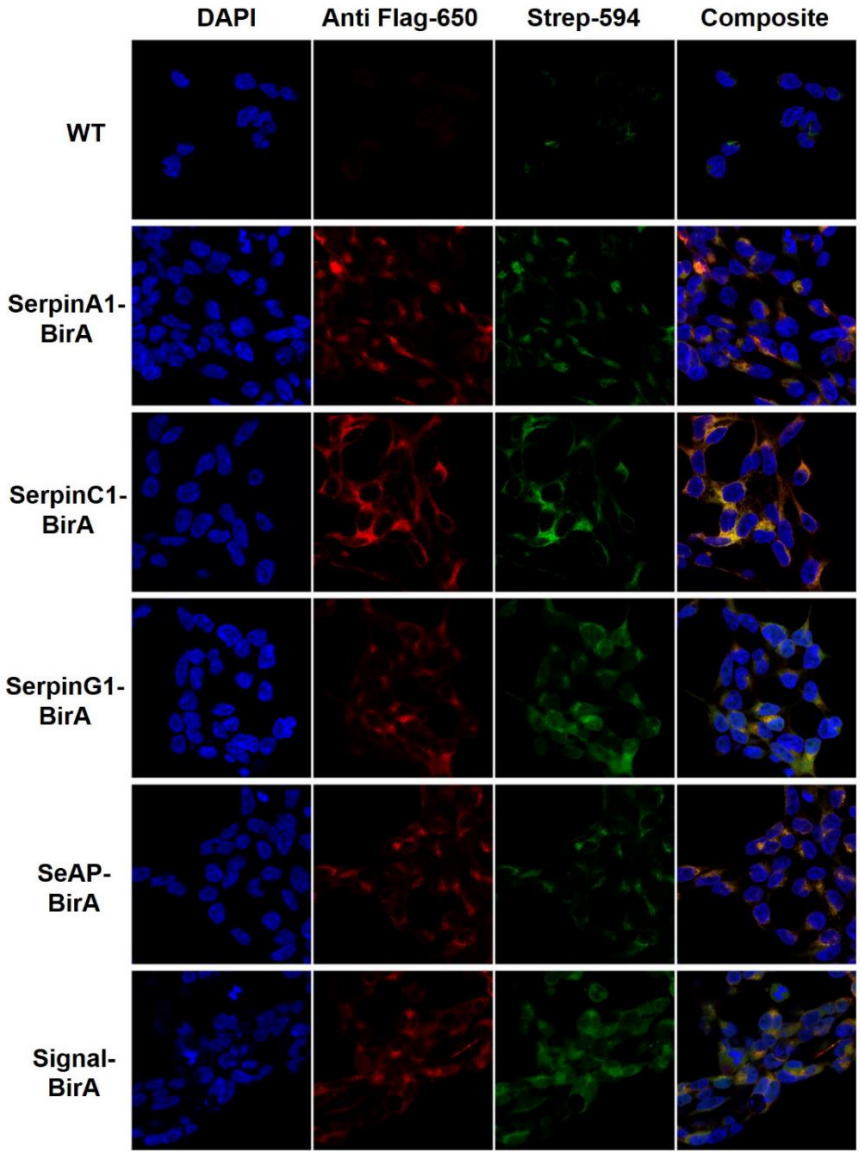


Figure 3. Bait-BirA fusion proteins are colocalized with biotin-staining. Co-Immunofluorescence demonstrated the intracellular colocalization of the biotin-labeled proteins (stained with Streptavidin-Dylight 594 and illustrated in green color) and bait-BirA (stained with anti-flag monoclonal antibody-Dylight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

To quantify colocalization of the bait-BirA fusion and biotinylated proteins, we calculated a range of Li's IQC from approximately 0.26 to 0.39 compared to the WT

calculated value of 0.22 from the co-immunofluorescence (Table 1 and Fig. S1), confirming the specificity of the BioID labeling system to tag the proximal proteins of SecPs (closer to 0.5 demonstrates a dependent protein staining pattern between the red and green channels).

Table 1. Colocalization metrics determined for each clone 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

WT sample revealed endogenous biotinylation landscape

While the BirA domains fused to the C-terminus of SEAP and the SERPIN family genes successfully biotinylated proximal proteins, we aimed to identify the interactions with each bait protein. Cells were then lysed to extract the total protein and biotinylated proteins purified using streptavidin. Proteins were digested by trypsin, and peptides were subjected to LC-MS/MS. MS spectra were then used to identify the biotinylated proteins in samples using MaxQuant[21], and differentially biotinylated proteins were identified in each sample compared to WT using Perseus[22]. When implemented in a WT control cell line (without a BirA fusion protein), we identified proteins that are biotinylated endogenously. These include proteins that bind biotin as a cofactor such as carboxylases (known to be problematic with streptavidin-based protein detection[23]). The endogenously biotinylated proteins remain after streptavidin-mediated purification, along with bona fide interactors. Although the extent to which more general endogenous biotinylation exists 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. Thus, using the interaction partners detected from the WT sample as background, we filtered out interactions detected in each model bait-BirA sample that were likely a result of endogenous biotinylation. After filtering endogenous biotinylated proteins, the top interactors were dominated by secretory pathway-resident and co-secreted proteins. Among the top differentially biotinylated proteins in bait-BirA samples, the bait proteins showed the highest log-fold change (LFC) (Fig. 4). 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 within the cell.

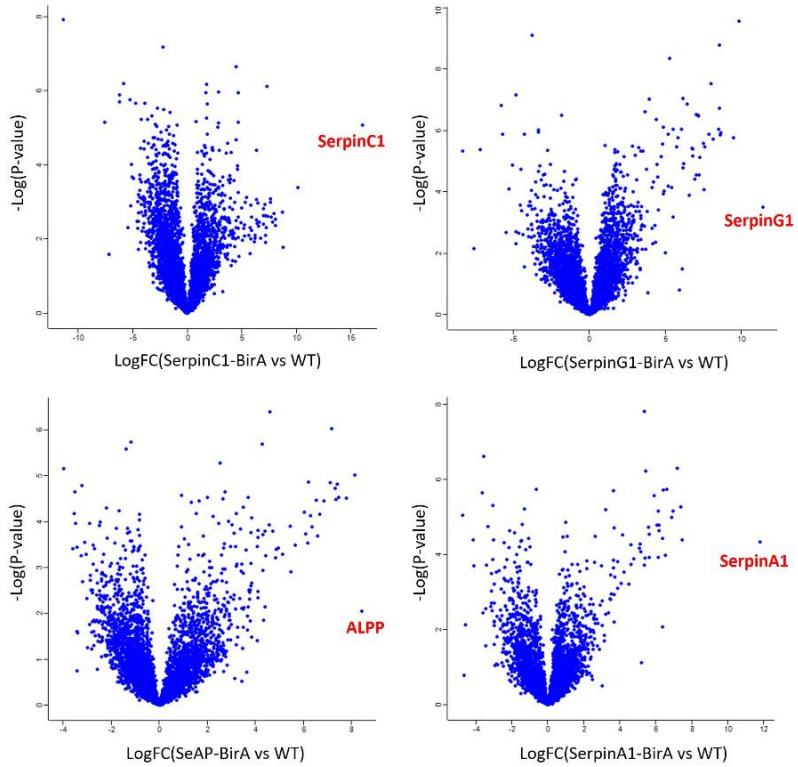


Figure 4. Dozens of proteins show significantly increased biotinylation after expression of bait-BirA proteins. Volcano plot showing the distribution of the quantified biotinylated proteins by MS according to p-value and fold change. As depicted the bait-protein significantly showed the highest fold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

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

For each secreted protein we identified probable PPIs as interactors having 3-fold 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) based on secretory pathway-related gold standards compiled from 3 independent gene sets (see methods). 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). 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 [24]. 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.

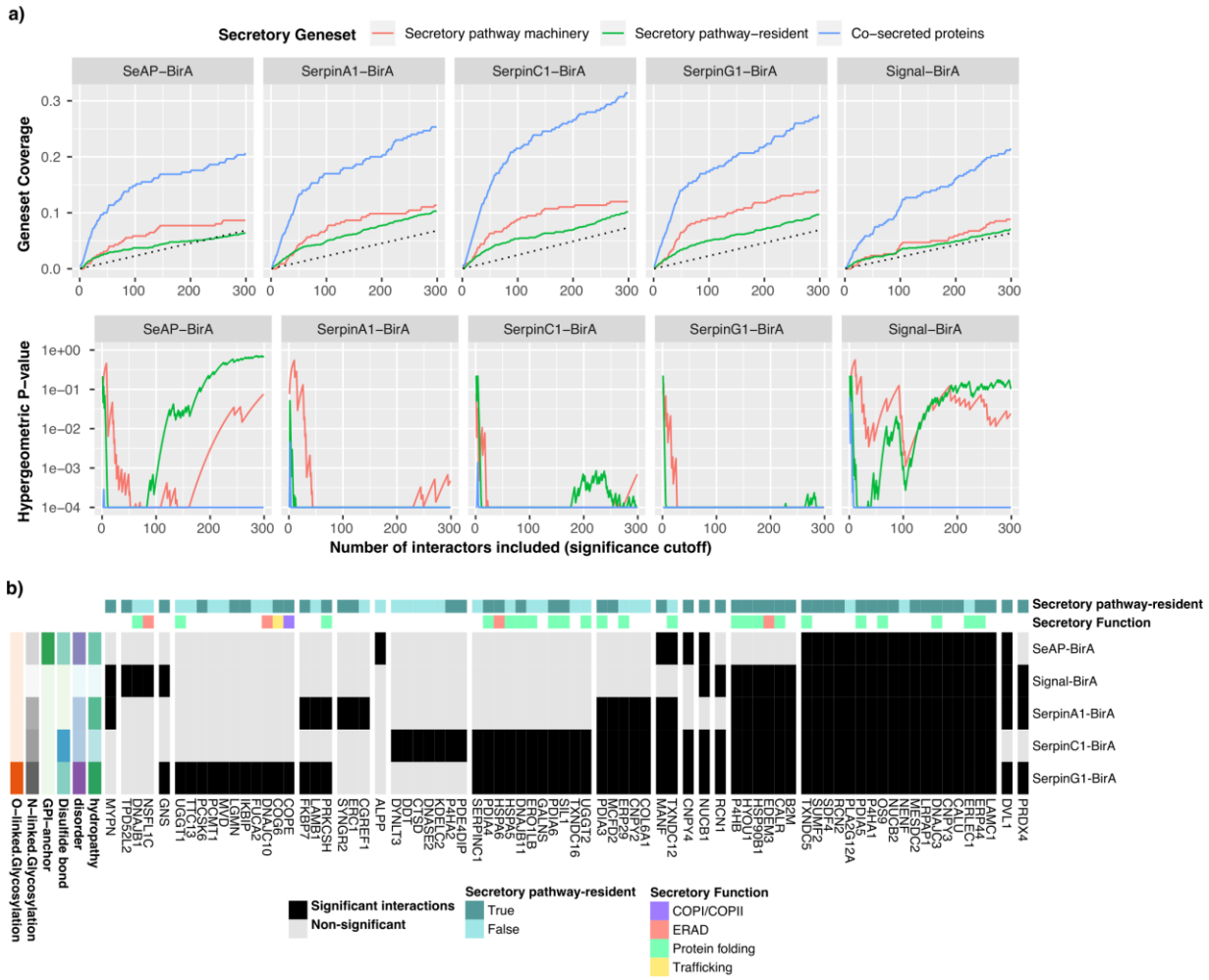


Figure 5. Interacting proteins are enriched for secretory pathway machinery. (a) To determine if significant interactions enrich for secretory pathway-related genes, we performed an iterative enrichment analysis in which we included the most significant interactions first and iteratively added interactors 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-values (bottom) are shown. 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 (x-axis) and the model proteins (y-axis), where the shadowed entries indicate significant interactions. The features of the model proteins, detailed in Fig. 6, are summarized for each model protein on the left and the interactors are labeled on top based on their secretory pathway attributes.

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

Protein-protein interactions mediate the folding, modification and transportation of secreted proteins [25–29]. 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[5]. The significant interactions captured between the bait-BirA proteins and their interactors link secreted

proteins to the responsible SecMs. As the bait-BirA proteins differ in structural composition and post-translational modifications (Fig. 6), 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 establish a more comprehensive connection around product-specific interactions with the secretory pathway, we aggregated various PTM and structural properties across all 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. 7a) that are known to help secretion of proteins with the corresponding PTMs. Thus, we looked more into the detected interactions associated with glycosylation, disulfide bond addition, and protein folding.

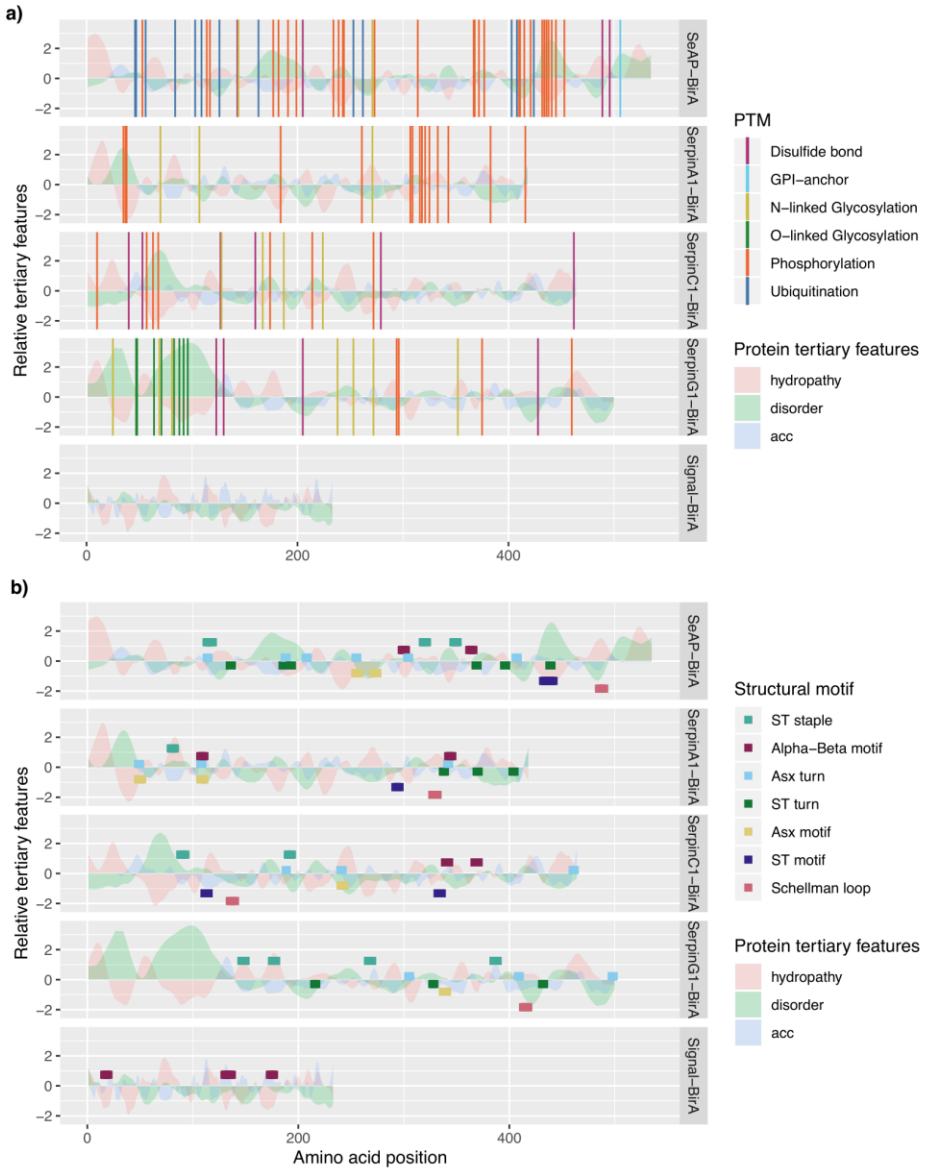


Figure 6. The bait proteins show diversity in their PTM and structural content. Dots and lines represent known PTM sites and structural motifs in panel (a) and (b) respectively. The hills and valleys indicate protein tertiary features. Note that solvent accessibility and structural motifs are only available for regions covered by the PDB structure, whereas predicted features such as protein hydrophobicity and disorder are available for the entire protein.

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. 7a). For example, the glycosylated bait proteins interacted with calreticulin (CALR), a calcium-binding chaperone that promotes folding, oligomeric assembly, and quality control of glycoproteins in the ER[30]. 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[31]. In addition, two members of the PDI family, PDIA3 and ERp29, which form a complex with calreticulin/calnexin, showed association with our N-linked glycan containing proteins (Fig. 7a) suggesting their role in glycoprotein folding and quality control. Calnexin/Calreticulin-PDIA3 complexes promote the oxidative folding of nascent polypeptides [32] and ERp29 promotes isomerization of peptidyl-prolyl bonds to attain the native polypeptide structure [32,33].

Proteins with chaperone activity, such as HSPA5 (Fig. 7a), were also found to interact with N-linked glycan-containing bait-BirA. HSPA5 is a chaperone that is a component of glycoprotein quality-control (GQC) system. GQC recognizes glycoproteins with amino acid substitutions, and targets them for ERAD[34]. EDEM3, another interactor associated with the N-glycan containing proteins (Fig. 7a), is a glycosyltransferase involved in ERAD mediated degradation of glycoproteins by catalyzing mannose trimming from Man8GlcNAc2 to Man7GlcNAc2 in N-glycans [35]. 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.

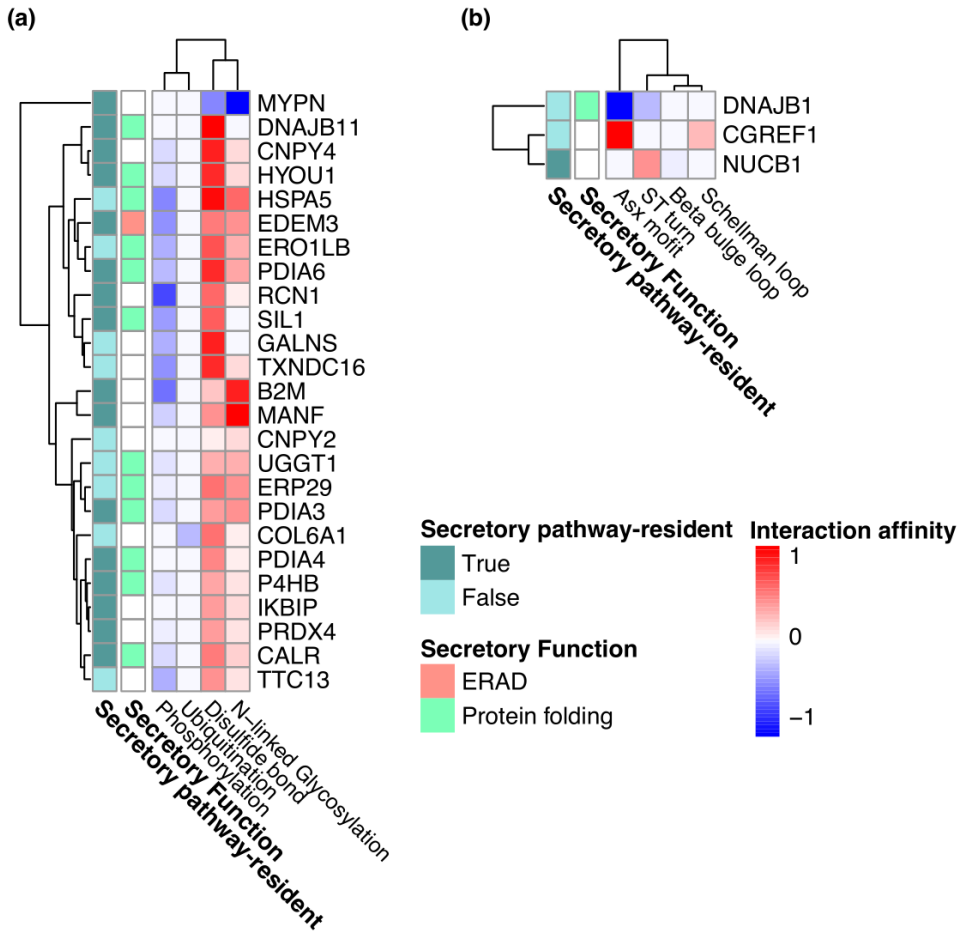


Figure 7. Detected interactors correlate with protein features. Interactors were associated with specific PTM (a) and structural (b) features of model proteins. The heatmap shows the standardized interaction affinities estimated between certain interactors and PTM and structural features across all model proteins (see methods). Only interactors having significant associations with model protein features are shown.

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. 7a). These enzymes catalyze the formation, breakage and rearrangement of disulfide bonds through the thioredoxin-like domains [36]. 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 the protein secretion. The proteins with more disulfide bonds also interact with major ER chaperones HSAP5 and DNAJB11, a co-chaperone of HSAP5, that play a key role in protein folding and quality control in the ER lumen [37,38], highlighting their important role in secretion of the disulfide bond enriched proteins.

The PDI, ERp44 showed the strongest association (LFC > 8, Fig. S4) with disulfide bond enriched proteins i.e. SERPINC1 and SERPING1. ERp44 mediates the ER localization of the oxidoreductase Ero1α (an oxidoreductin that reoxidizes P4HB to enable additional rounds of disulfide formation) through the formation of reversible mixed

disulfides[39]. Hence, the strong association of the ERp44 suggests the importance of the thiol-mediated retention in disulfide bond formation, particularly when secretory is loaded with the proteins with dominant disulfide bond. In addition, ERO1LB, PRDX4 and SIL1 were ER-localized enzymes that were associated with disulfide bond formation. ERO1LB efficiently reoxidizes P4HB [40], PRDX4 couples hydrogen peroxide catabolism with oxidative protein folding by reducing hydrogen peroxide [41], and SIL1 can reverse HSAP5 cysteine oxidation which alters its chaperone activity to cope with suboptimal folding conditions [42]. 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 the ERp44 interaction, which showed the highest fold change, on productivity of proteins with more disulfide bonds we knocked down ERP44 in the cells expressing SERPINC1-BirA, using an orthogonal RNAi approach, esiRNAs, to target gene knockdown [43]. HEK293 cells expressing SERPINC1-BirA were transfected with esiRNA against ERP44, with EGFP and KIF11 as negative/positive controls, respectively. The supernatants of esiRNA experiments collected 48- and 72-hours post-transfection for measuring the secretion of SERPINC1-BirA by ELISA. All quantifications were repeated in duplicate. The results showed ERP44 knockdown in SERPINC1-BirA cells led to the 44% and 41% reduction in the secretion of SERPINC1 at day two and three post-transfection, respectively, compared to the negative control, supporting the importance of the Thiol-mediated retention on secretion of the disulfide bond enriched protein.

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. 7b). For example, model proteins depleted in the asx motif [44] showed higher tendency to interact with DNAJB1, a molecular chaperone of the HSP70 family. 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 [45,46]. As another example, NUCB1, a chaperone-like amyloid binding protein that prevents protein aggregation [47], interacted more strongly with our proteins with more ST-turns (Fig. 7b). ST turns occur frequently at the N-termini of α-helices as part of ST motifs[48] and are regarded as helix capping features which stabilize α-helices in proteins [49]. 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 [14]. However, this is the first time that BioID has been used to identify the required supporting machinery of recombinant therapeutic proteins, while they

are being synthesized and trafficked along the secretory pathway. Numerous protein-protein interactions guiding the folding, modifications, and trafficking of the secreted and membrane proteins through the secretory pathway are transient [50,51] and therefore cannot be captured by the conventional methods such as co-immunoprecipitation. Consistent with the previous studies [52], these results showed the BioID can detect weak and transient interactions *in situ.* This is important since numerous interactions in the secretory pathway are transient (e.g., chaperones aiding folding or enzymes adding PTMs). Thus, it is a powerful approach to study luminal processes involved in protein secretion.

We found that the disulfide bridge formation enzymes showed the most significant association with bait proteins most enriched in disulfide bonds, supporting their critical roles in protein secretion and maintaining the ER stability. Insufficient interaction between secreted proteins and PDIs can limit secretion efficiency and serve as a rate-limiting step for secretion of proteins with enriched disulfide bonds. Our results are consistent with a previous study on difficult to express (DTE) monoclonal antibodies wherein the disulfide bridge formation within the antibody light chain (LC) was impaired due to less recognition by PDI. Therefore, the DTE molecule was degraded intracellularly by the ubiquitin proteasome system via ER-associated degradation (ERAD) [55]. In another study, the tissue specific analysis of SecM 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 [5]. P4HB and PDIA4 catalyze the formation, breakage and rearrangement of disulfide bonds. These observations are clear evidence which 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 which 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 to the secretion of those proteins. These interactions can potentially be a bottleneck especially if they are expressed lowly within the cell. While evidence linking SecMs to motif-specific modification is lacking, many molecular chaperones selectively interact with certain sequence and structural elements to favor the particular folding pathways[56]. 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. With the emergence of DTE therapeutics that have complex biological structures, special attention is therefore needed for optimizing cellular synthesis in a client specific manner.

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 mammalian 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 [53]. Here we showed this challenge is not an insurmountable issue, in that the BioID2 construct successfully profiled the interactome of the secretory pathway, and we identified many luminal interactions with expected SecM components. BioID2 requires less biotin supplementation, and exhibits enhanced labeling of proximate protein [20]. BioID2 was shown to improve localization and was more sensitive to a lower biotin concentration, potentially allowing for BioID to be introduced to new systems where biotinylation supplementation may not be easily accomplished [52]. More recently, Ting's group has developed two promiscuous mutants of biotin ligase, TurboID and miniTurbo, which catalyze proximity labeling even with much greater efficiency [54]. Therefore, BioID2 or TurboID can be considered an effective method when proximal labeling of the endomembrane organelles is desired.

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 up 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.

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 followed by 3XFLAG tag to simplify the immuno detection. For constitutive expression, 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. 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). Using Gibson Assembly Master Mix (NEB) DNA fragments were assembled into the vector according to the manufacturer's instructions, assembly products were transformed to the chemically competent *e. coli*, and recombinant plasmids were extracted for further validation. Successful cloning of the constructs was verified by restriction digestion followed by gel electrophoresis, and sequencing. All constructs had a glycine-serine linker added between BirA and the model proteins to provide flexibility between them as well as 3X FLAG tag to ease the further detection of proteins. 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 uM biotin on acid-washed 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. Co-immunofluorescence staining was then performed to study the intracellular colocalization of bait-BirA proteins and biotinylated interactors. For this, flag tag 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 of stained bait-BirA and biotin using Fiji's (ImageJ 1.52p) Coloc_2 analysis tool between the 650 (Anti-flag) and 594 (Streptavidin) channels [57]. 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) [58,59]. The Pearson's and Manders' colocalization coefficients are used to determine the degree of overlap between images and are mathematically similar; however, the Pearson's coefficient utilizes deviation from the average intensity of each channel, whereas the MCC are each relative to the total fluorescence in each channel [51]. The MCC are said to allow for strong differences between channel signal intensities due to labeling or settings during/after image capture, however, it is more susceptible to offset and background where the Person's value is not. Li's IQC is also derived from Pearson's coefficient, but further aids in determining the dependency of staining intensities, i.e. pixel channel intensities vary around mean channel intensities together when there is synchrony of two proteins in a complex [50]. 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.

Western blotting

To validate the secretion of bait-BirA proteins, supernatants of cultures expressing fusion proteins were collected, centrifuged to remove cell debris, and protein content was concentrated using Amicon Ultra 4 mL 10 KD filter unit (MilliporeSigma) and quantified using Bradford assay (Expedeon). 20 ug of total protein was loaded on SDS-PAGE gel for electrophoresis and resolved proteins were transblotted to nitrocellulose membrane using 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 tag 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 and 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.

ELISA

Aliquots of clarified supernatants from esiRNA transfected cultures were taken out of −80°C and thawn on the ice and immediately subjected to ELISA. FLAG levels were quantified from secreted bait-BirA fusion proteins by competitive anti-flag ELISA using the DYKDDDDK-Tag Detection ELISA Kit (Cayman Chemicals, catalog # 501560). All measurements were performed according to the manufacture instructions in triplicate. The effect of PDIA's knockdown on secretion of the model proteins was measured in comparison with the negative control of each cell line transfected with EGFP esiRNA as negative control (see below).

RNAi knockdown experiment

esiRNA technology developed by Eupheria Biotech was applied for knocking candidate PDIAs genes because it already showed highly specific and effective gene knockdowns with lower off-target effects than single, chemically synthesized siRNA. esiRNA targeting ERP44 was ordered from Sigma. HEK293 cell expressing SERPINC1-BirA were seeded at $0.6X10^5$ cells/well in 24-well plates with complete medium and reverse transfected with 72 ng of ERP44 specific esiRNA or EGFP 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. Targeted gene knockdown by esiRNA was allowed to occur for 48 and 72 h, culture supernatants were harvested, clarified by low-speed centrifugation, then aliquoted and stored at -80 for further experiments.

Sample preparation for biotinylation

Cells stably expressing bait-BirA (and suitable controls) were grown in 245 mm plates (one plate per biological replicate in triplicate) to approximately 70% confluence in complete media. 50 µM biotin (50 µL of a 20 mM stock per 20 mL media) was then added to the culture and incubated with the cells for 24 h. The next day the medium was removed, and the cells were then rinsed gently first with 60 mL of room temperature PBS and then washed with ice-cold PBS to remove residual amounts of serum from cells. Cell were scraped from tissue culture vessel in PBS, resuspended by repeated aspirations, and transferred to clean tube to be centrifuged at a maximum of 500×g for 5 min at 4 C. Cells were washed once again in cold PBS and pellets were frozen on dry ice and stored at −80°C until ready to process.

Mass Spectrometry

Cells pellets were 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). Briefly, cartridges were first primed with lysis buffer. Samples were then loaded onto a streptavidin cartridge and background contamination was removed by washing the cartridges with lysis buffer followed by rapid digestion buffer (Promega, Rapid digestion buffer kit), 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 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 [21] version 1.5.5.1. MS/MS spectra were searched against the Homo sapiens Uniport 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 from BioID-tagged stable cell lines relative to the WT cell line 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 [60] integrated into Normalyzer [61]. Perseus software [22] was employed for data preparation, filtering, and computation of differential protein abundance. The DEP package[62] was used to explore the pattern of the missing values and check 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 $\log_2$ transformed data.

Detection of significant interactions

The threshold for significant interactions was determined using the known secretory pathway components as a gold standard. More specifically, 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 that the optimal cutoffs are robust to the gold standards chosen.

Aggregation of interactions and estimation of interaction synergy

To further generalize the interactions between individual SecPs and their interactors, we aggregated proteins based on several features. 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. These common properties include shared structural motifs [63], known sites of PTM from Uniport and phosphosite [63,64]. The secretory pathway interactors can also be aggregated into curated biological pathways (cite mSigDB[65], KEGG[66], Reactome[67]) and subcellular localizations[68].

Bayesian modeling framework

To quantify such associations of each interacting protein with protein features, we calculate 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 SecP p (f_p) .

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

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

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, $\delta_{f,g}$, takes 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. We modeled the interaction frequency $\delta_{f,g}$ with a Poisson distribution.

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

Interaction effects:

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

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 feature promiscuity (number of SecMs g connected to a feature), interactor promiscuity (number of SecPs with a feature f interacting with a SecM) and an intercept variable. The residual becomes the interaction synergy between *f* and *g,* quantifying the degree to which f and g interact more than by random chance. In previous work, this approach has correctly estimated

epistasis intensity[69]. To better regularize the synergies and promiscuity, their Bayesian priors are all normally distributed around 0.

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). The authors would also like to thank Tune Wulff for support on pilot assays.

Conflict of Interest

The authors declare no competing interests.

References

(1) Global Plasma Protein Therapeutics Market Analysis - By Product Type (Immunoglobulins, Albumin, Coagulation Factors, Alpha-1 Proteinase Inhibitor, Others) By Application (Hemophilia, Primary Immunodeficiency Disorder, Idiopathic Thrombocytopenic Purpura, Secondary Immunodeficiency And Others), By Region - Industry Size, Share, Growth, Trends and Forecast | 2019 - 2024 https://www.marketdataforecast.com/market-reports/global-plasma-protein-therapeutics-market.

(2) Matasci, M.; Hacker, D. L.; Baldi, L.; Wurm, F. M. Recombinant Therapeutic Protein Production in Cultivated Mammalian Cells: Current Status and Future Prospects. *Drug Discov. Today Technol.* **2008**, *5* (2-3), e37–e42.

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

(4) Gutierrez, J. M.; Feizi, A.; Li, S.; Kallehauge, T. B.; Hefzi, H.; Grav, L. M.; Ley, D.; Baycin Hizal, D.; Betenbaugh, M. J.; Voldborg, B.; et al. Genome-Scale Reconstructions of the Mammalian Secretory Pathway Predict Metabolic Costs and Limitations of Protein Secretion. *Nat. Commun.* **2020**, *11* (1), 68.

(5) Feizi, A.; Gatto, F.; Uhlen, M.; Nielsen, J. Human Protein Secretory Pathway Genes Are Expressed in a Tissue-Specific Pattern to Match Processing Demands of the Secretome. *npj Systems Biology and Applications*. 2017. https://doi.org/10.1038/s41540-017-0021-4.

(6) Lund, A. M.; Kaas, C. S.; Brandl, J.; Pedersen, L. E.; Kildegaard, H. F.; Kristensen, C.; Andersen, M. R. Network Reconstruction of the Mouse Secretory Pathway Applied on CHO Cell Transcriptome Data. *BMC Systems Biology*. 2017. https://doi.org/10.1186/s12918-017-0414-4.

(7) Novick, P.; Ferro, S.; Schekman, R. Order of Events in the Yeast Secretory Pathway. *Cell* **1981**, *25* (2), 461–469.

(8) Reynaud, E. G.; Simpson, J. C. Navigating the Secretory Pathway: Conference on Exocytosis Membrane Structure and Dynamics. *EMBO Rep.* **2002**, *3* (9), 828–833.

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

(10) Hussain, H.; Maldonado-Agurto, R.; Dickson, A. J. The Endoplasmic Reticulum and Unfolded Protein Response in the Control of Mammalian Recombinant Protein Production. *Biotechnol. Lett.* **2014**, *36* (8), 1581–1593.

(11) Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells. *J. Cell Biol.* **2012**, *196* (6), 801–810.

(12) Kim, D. I.; Birendra, K. C.; Zhu, W.; Motamedchaboki, K.; Doye, V.; Roux, K. J. Probing Nuclear Pore Complex Architecture with Proximity-Dependent Biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (24), E2453–E2461.

(13) Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **2013**, *339* (6125), 1328–1331.

(14) Varnaitė, R.; MacNeill, S. A. Meet the Neighbors: Mapping Local Protein Interactomes by Proximity-Dependent Labeling with BioID. *Proteomics* **2016**, *16* (19), 2503–2518.

(15) Christopher D. Go, James D.R. Knight, Archita Rajasekharan, Bhavisha Rathod, Geoffrey G. Hesketh, Kento T. Abe, Ji-Young Youn, Payman Samavarchi-Tehrani, Hui Zhang, Lucie Y. Zhu, Evelyn Popiel, Jean-Philippe Lambert, Étienne Coyaud, Sally W.T. Cheung, Dushyandi Rajendran, Cassandra J. Wong, Hana Antonicka, Laurence Pelletier, Brian Raught, Alexander F. Palazzo, Eric A. Shoubridge, Anne-Claude Gingras. A Proximity Biotinylation Map of a Human Cell. *bioRxiv* **2019**. https://doi.org/10.1101/796391.

(16) Firat-Karalar, E. N.; Stearns, T. Probing Mammalian Centrosome Structure Using BioID Proximity-Dependent Biotinylation. *Methods Cell Biol.* **2015**, *129*, 153–170.

(17) Gupta, G. D.; Coyaud, É.; Gonçalves, J.; Mojarad, B. A.; Liu, Y.; Wu, Q.; Gheiratmand, L.; Comartin, D.; Tkach, J. M.; Cheung, S. W. T.; et al. A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **2015**, *163* (6), 1484–1499.

(18) Dong, J.-M.; Tay, F. P.-L.; Swa, H. L.-F.; Gunaratne, J.; Leung, T.; Burke, B.; Manser, E. Proximity Biotinylation Provides Insight into the Molecular Composition of Focal Adhesions at the Nanometer Scale. *Sci. Signal.* **2016**, *9* (432), rs4.

(19) Hoffman, A. M.; Chen, Q.; Zheng, T.; Nicchitta, C. V. Heterogeneous Translational Landscape of the Endoplasmic Reticulum Revealed by Ribosome Proximity Labeling and Transcriptome Analysis. *J. Biol. Chem.* **2019**, *294* (22), 8942–8958.

(20) Kim, D. I.; Jensen, S. C.; Noble, K. A.; Kc, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J. An Improved Smaller Biotin Ligase for BioID Proximity Labeling. *Mol. Biol. Cell* **2016**, *27* (8), 1188–1196.

(21) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319.

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

(23) Tytgat, H. L. P.; Schoofs, G.; Driesen, M.; Proost, P.; Van Damme, E. J. M.; Vanderleyden, J.; Lebeer, S. Endogenous Biotin-Binding Proteins: An Overlooked Factor Causing False Positives in Streptavidin-Based Protein Detection. *Microb. Biotechnol.* **2015**, *8* (1), 164–168.

(24) Mayer, M. P. Gymnastics of Molecular Chaperones. *Mol. Cell* **2010**, *39* (3), 321–331.

(25) Calakos, N.; Bennett, M. K.; Peterson, K. E.; Scheller, R. H. Protein-Protein Interactions Contributing to the Specificity of Intracellular Vesicular Trafficking. *Science* **1994**, *263* (5150), 1146–1149.

(26) Pearl, L. H.; Prodromou, C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. *Annu. Rev. Biochem.* **2006**, *75*, 271–294.

(27) Watanabe, S.; Amagai, Y.; Sannino, S.; Tempio, T.; Anelli, T.; Harayama, M.; Masui, S.; Sorrentino, I.; Yamada, M.; Sitia, R.; et al. Zinc Regulates ERp44-Dependent Protein Quality Control in the Early Secretory Pathway. *Nat. Commun.* **2019**, *10* (1), 603.

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

(29) Ikawa, M.; Wada, I.; Kominami, K.; Watanabe, D.; Toshimori, K.; Nishimune, Y.; Okabe, M. The Putative Chaperone Calmegin Is Required for Sperm Fertility. *Nature* **1997**, *387* (6633), 607–611.

(30) Nauseef, W. M.; McCormick, S. J.; Clark, R. A. Calreticulin Functions as a Molecular Chaperone in the Biosynthesis of Myeloperoxidase. *J. Biol. Chem.* **1995**, *270* (9), 4741–4747.

(31) Ferris, S. P.; Jaber, N. S.; Molinari, M.; Arvan, P.; Kaufman, R. J. UDP-Glucose:glycoprotein Glucosyltransferase (UGGT1) Promotes Substrate Solubility in the Endoplasmic Reticulum. *Mol. Biol. Cell* **2013**, *24* (17), 2597–2608.

(32) Sakono, M.; Seko, A.; Takeda, Y.; Ito, Y. PDI Family Protein ERp29 Forms 1:1 Complex with Lectin Chaperone Calreticulin. *Biochem. Biophys. Res. Commun.* **2014**, *452* (1), 27–31.

(33) Tannous, A.; Pisoni, G. B.; Hebert, D. N.; Molinari, M. N-Linked Sugar-Regulated Protein Folding and Quality Control in the ER. *Semin. Cell Dev. Biol.* **2015**, *41*, 79–89.

(34) Ferris, S. P.; Kodali, V. K.; Kaufman, R. J. Glycoprotein Folding and Quality-Control Mechanisms in Protein-Folding Diseases. *Dis. Model. Mech.* **2014**, *7* (3), 331–341.

(35) Ninagawa, S.; Okada, T.; Sumitomo, Y.; Kamiya, Y.; Kato, K.; Horimoto, S.; Ishikawa, T.; Takeda, S.; Sakuma, T.; Yamamoto, T.; et al. EDEM2 Initiates Mammalian Glycoprotein ERAD by Catalyzing the First Mannose Trimming Step. *J. Cell Biol.* **2014**, *206* (3), 347–356.

(36) Kozlov, G.; Määttänen, P.; Thomas, D. Y.; Gehring, K. A Structural Overview of the PDI Family of Proteins. *FEBS J.* **2010**, *277* (19), 3924–3936.

(37) Ng, D. T.; Watowich, S. S.; Lamb, R. A. Analysis in Vivo of GRP78-BiP/substrate Interactions and Their Role in Induction of the GRP78-BiP Gene. *Mol. Biol. Cell* **1992**, *3* (2), 143–155.

(38) Yu, M.; Haslam, R. H.; Haslam, D. B. HEDJ, an Hsp40 Co-Chaperone Localized to the Endoplasmic Reticulum of Human Cells. *J. Biol. Chem.* **2000**, *275* (32), 24984–24992.

(39) Anelli, T.; Alessio, M.; Bachi, A.; Bergamelli, L.; Bertoli, G.; Camerini, S.; Mezghrani, A.; Ruffato, E.; Simmen, T.; Sitia, R. Thiol-Mediated Protein Retention in the Endoplasmic Reticulum: The Role of ERp44. *EMBO J.* **2003**, *22* (19), 5015–5022.

(40) Mezghrani, A.; Fassio, A.; Benham, A.; Simmen, T.; Braakman, I.; Sitia, R. Manipulation of Oxidative Protein Folding and PDI Redox State in Mammalian Cells. *EMBO J.* **2001**, *20* (22), 6288–6296.

(41) Zito, E. PRDX4, an Endoplasmic Reticulum-Localized Peroxiredoxin at the Crossroads between Enzymatic Oxidative Protein Folding and Nonenzymatic Protein Oxidation. *Antioxid. Redox Signal.* **2013**, *18* (13), 1666–1674.

(42) Siegenthaler, K. D.; Pareja, K. A.; Wang, J.; Sevier, C. S. An Unexpected Role for the Yeast Nucleotide Exchange Factor Sil1 as a Reductant Acting on the Molecular Chaperone BiP. *Elife* **2017**, *6*. https://doi.org/10.7554/eLife.24141.

(43) Kittler, R.; Heninger, A.-K.; Franke, K.; Habermann, B.; Buchholz, F. Production of Endoribonuclease-Prepared Short Interfering RNAs for Gene Silencing in Mammalian Cells. *Nat. Methods* **2005**, *2* (10), 779–784.

(44) Golovin, A.; Henrick, K. MSDmotif: Exploring Protein Sites and Motifs. *BMC Bioinformatics* **2008**, *9*, 312.

(45) Imperiali, B.; Shannon, K. L.; Rickert, K. W. Role of Peptide Conformation in Asparagine-Linked Glycosylation. *Journal of the American Chemical Society*. 1992, pp 7942–7944. https://doi.org/10.1021/ja00046a068.

(46) Imperiali, B.; Shannon, K. L.; Unno, M.; Rickert, K. W. Mechanistic Proposal for Asparagine-Linked Glycosylation. *Journal of the American Chemical Society*. 1992, pp 7944–7945. https://doi.org/10.1021/ja00046a069.

(47) Bonito-Oliva, A.; Barbash, S.; Sakmar, T. P.; Graham, W. V. Nucleobindin 1 Binds to Multiple Types of Pre-Fibrillar Amyloid and Inhibits Fibrillization. *Sci. Rep.* **2017**, *7*, 42880.

(48) Doig, A. J.; Stapley, B. J.; Macarthur, M. W.; Thornton, J. M. Structures of N-Termini of Helices in Proteins. *Protein Science*. 2008, pp 147–155. https://doi.org/10.1002/pro.5560060117.

(49) Aurora, R.; Rosee, G. D. Helix Capping. *Protein Science*. 1998, pp 21–38. https://doi.org/10.1002/pro.5560070103.
(50) Nyfeler, B.; Michnick, S. W.; Hauri, H.-P. Capturing Protein Interactions in the Secretory Pathway of Living Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (18), 6350–6355.
(51) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* **2009**, *109* (3), 839–860.
(52) Sears, R. M.; May, D. G.; Roux, K. J. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol. Biol.* **2019**, *2012*, 299–313.
(53) Kim, D. I.; Roux, K. J. Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends Cell Biol.* **2016**, *26* (11), 804–817.
(54) Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Author Correction: Efficient Proximity Labeling in Living Cells and Organisms with TurboID. *Nat. Biotechnol.* **2020**, *38* (1), 108.
(55) Mathias, S.; Wippermann, A.; Raab, N.; Zeh, N.; Handrick, R.; Gorr, I.; Schulz, P.; Fischer, S.; Gamer, M.; Otte, K. Unraveling What Makes a Monoclonal Antibody Difficult-to-express: From Intracellular Accumulation to Incomplete Folding and Degradation via ERAD. *Biotechnology and Bioengineering*. 2020, pp 5–16. https://doi.org/10.1002/bit.27196.
(56) Gidalevitz, T.; Stevens, F.; Argon, Y. Orchestration of Secretory Protein Folding by ER Chaperones. *Biochim. Biophys. Acta* **2013**, *1833* (11), 2410–2424.
(57) Schindelin, J.; Rueden, C. T.; Hiner, M. C.; Eliceiri, K. W. The ImageJ Ecosystem: An Open Platform for Biomedical Image Analysis. *Mol. Reprod. Dev.* **2015**, *82* (7-8), 518–529.
(58) Li, Q. A Syntaxin 1, G O, and N-Type Calcium Channel Complex at a Presynaptic Nerve Terminal: Analysis by Quantitative Immunocolocalization. *Journal of Neuroscience*. 2004, pp 4070–4081. https://doi.org/10.1523/jneurosci.0346-04.2004.
(59) Manders, E. M. M.; Verbeek, F. J.; Aten, J. A. Measurement of Co-Localization of Objects in Dual-Colour Confocal Images. *Journal of Microscopy*. 1993, pp 375–382. https://doi.org/10.1111/j.1365-2818.1993.tb03313.x.
(60) Smyth, G. K. 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.
(61) Chawade, A.; Alexandersson, E.; Levander, F. Normalyzer: A Tool for Rapid Evaluation of Normalization Methods for Omics Data Sets. *J. Proteome Res.* **2014**, *13* (6), 3114–3120.
(62) Zhang, X.; Smits, A. H.; van Tilburg, G. B.; Ovaa, H.; Huber, W.; Vermeulen, M. Proteome-Wide Identification of Ubiquitin Interactions Using UbIA-MS. *Nat. Protoc.* **2018**, *13* (3), 530–550.
(63) Blatch, G. L.; Lässle, M. The Tetratricopeptide Repeat: A Structural Motif Mediating Protein-Protein Interactions. *Bioessays* **1999**, *21* (11), 932–939.
(64) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (Database issue), D512–D520.
(65) Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdóttir, H.; Tamayo, P.; Mesirov, J. P. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* **2011**, *27* (12), 1739–1740.
(66) Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* **2017**, *45* (D1), D353–D361.
(67) Croft, D.; Mundo, A. F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M. R.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2014**, *42* (Database issue), D472–D477.
(68) Thul, P. J.; Åkesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm, T.; Asplund, A.; Björk, L.; Breckels, L. M.; et al. A Subcellular Map of the Human Proteome. *Science* **2017**, *356* (6340). https://doi.org/10.1126/science.aal3321.

(69) Shen, J. P.; Zhao, D.; Sasik, R.; Luebeck, J.; Birmingham, A.; Bojorquez-Gomez, A.; Licon, K.; Klepper, K.; Pekin, D.; Beckett, A. N.; et al. Combinatorial CRISPR-Cas9 Screens for de Novo Mapping of Genetic Interactions. *Nat. Methods* **2017**, *14* (6), 573–576.