Published in final edited form as:

J Proteome Res. 2015 November 6; 14(11): 4687–4703. doi:10.1021/acs.jproteome.5b00588.

Elucidation of the CHO Super-Ome (CHO-SO) by Proteoinformatics

Amit Kumar^{†,‡}, Deniz Baycin-Hizal[▽], Daniel Wolozny[†], Lasse Ebdrup Pedersen[§], Nathan E. Lewis^{⊥,¶}, Kelley Heffner[†], Raghothama Chaerkady[#], Robert N. Cole[#], Joseph Shiloach[‡], Hui Zhang^{||}, Michael A. Bowen[▽], and Michael J. Betenbaugh^{*,†}

[†]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States

[‡]Biotechnology Core Laboratory, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health, Building 14A, Bethesda, Maryland 20892, United States

[§]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark

[⊥]Department of Biology, Brigham Young University, Provo, Utah 84602, United States

[¶]Department of Pediatrics, University of California, San Diego, California 92093, United States

[#]Institute of Basic Biomedical Sciences, Mass Spectrometry and Proteomics Facility, Johns Hopkins University School of Medicine, 733 North Broadway Street, Baltimore, Maryland 21205, United States

^{||}Department of Pathology, Johns Hopkins School of Medicine, 400 North Broadway Street, Baltimore, Maryland 21287, United States

[▽]Antibody Discovery and Protein Engineering, MedImmune LLC, 1 MedImmune Way, Gaithersburg, Maryland 20878, United States

Abstract

*Corresponding Author: beten@jhu.edu. Phone: +1 410-516-5461. Fax: +1 410-516-5510.

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteo-me.5b00588. Overview of contents; complete overview of all proteins; detailed results from search engines used for all 3281 identified proteins; detailed results from search engines used for potentially secreted proteins; final filtered data set of CHO-SO; GO analyses of molecular functions, biological processes, and cellular components; NSAF values of all 3281 identified proteins; high-abundance proteins that are potentially HCPs; immunogenicity results (XLSX)

Chinese hamster ovary (CHO) cells are the preferred host cell line for manufacturing a variety of complex biotherapeutic drugs including monoclonal antibodies. We performed a proteomics and bioinformatics analysis on the spent medium from adherent CHO cells. Supernatant from CHO-K1 culture was collected and subjected to in-solution digestion followed by LC/LC-MS/MS analysis, which allowed the identification of 3281 different host cell proteins (HCPs). To functionally categorize them, we applied multiple bioinformatics tools to the proteins identified in our study including SignalP, TargetP, SecretomeP, TMHMM, WoLF PSORT, and Phobius. This analysis provided information on the presence of signal peptides, transmembrane domains, and cellular localization and showed that both secreted and intracellular proteins were constituents of the supernatant. Identified proteins were shown to be localized to the secretory pathway including ones playing roles in cell growth, proliferation, and folding as well as those involved in protein degradation and removal. After combining proteins predicted to be secreted or having a signal peptide, we identified 1015 proteins, which we termed as CHO supernatant-ome (CHO-SO), or superome. As a part of this effort, we created a publically accessible web-based tool called GO-CHO to functionally categorize proteins found in CHO-SO and identify enriched molecular functions, biological processes, and cellular components. We also used a tool to evaluate the immunogenicity potential of high-abundance HCPs. Among enriched functions were catalytic activity and structural constituents of the cytoskeleton. Various transport related biological processes, such as vesicle mediated transport, were found to be highly enriched. Extracellular space and vesicular exosome associated proteins were found to be the most enriched cellular components. The superome also contained proteins secreted from both classical and nonclassical secretory pathways. The work and database described in our study will enable the CHO community to rapidly identify high-abundance HCPs in their cultures and therefore help assess process and purification methods used in the production of biologic drugs.

Keywords

proteomics; secretome; CHO; signal peptides; ontology; host cell proteins; immunogenicity

1. INTRODUCTION

Mammalian cell lines are the preferred hosts for the production of numerous recombinant proteins, especially those of biotherapeutic interest, due to their ability to secrete complex proteins with post-translational modifications compatible for use as drugs in humans. Because of the ability of mammalian cells to synthesize and secrete complex proteins, 32 biotherapeutics products from mammalian cells were approved by regulatory authorities between 2006 and 2010,¹ and current trends project that monoclonal antibody production in mammalian-based system will double in value between 2010 and 2016.² Among mammalian cells, Chinese hamster ovary (CHO) cells have been the most widely used cell line due to their ease of cultivation in suspension and adaptability to different media compositions, both critical factors in high level protein production in large volume bioreactors. To study these cellular factories, proteomics can serve as a useful tool to identify and quantify secreted proteins and to provide general insights into physiology of cellular proteins. This information may lead to further improvement of the CHO cells' production capabilities and assist in eliminating undesirable proteins during the purification

process. CHO cell proteomics has previously been used to identify and quantify the proteins involved in growth, protein bioprocessing, metabolism, glycosylation, and apoptosis.³ However, there has only been a limited analysis of extracellular or secreted CHO proteins.⁴

By using mass spectrometry (MS)-based proteomics methods, we have characterized the supernatant of CHO cell culture postcentrifugation. The protein groups identified in this supernatant may influence cell growth, development, differentiation, and many other cell features. Computational analysis was used to characterize the CHO supernatant-ome (CHO-SO) and to filter potentially secreted proteins using different bioinformatics tools including (1) SignalP (<http://www.cbs.dtu.dk/services/SignalP/>), which predicts proteins secreted by the classical pathway, (2) SecretomeP (<http://www.cbs.dtu.dk/services/SecretomeP/>), which predicts proteins secreted by the nonclassical pathways,⁵ (3) TMHMM (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), which predicts transmembrane helices,⁶ (4) Phobius (<http://phobius.sbc.su.se/>), which predicts signal peptides and various regions of a transmembrane protein sequence,⁷ (5) TargetP (<http://www.cbs.dtu.dk/services/TargetP/>), which predicts signal peptides in a protein sequences as well as their subcellular locations,⁸ (6) WoLF PSORT (<http://www.genscript.com/wolf-psort.html>), which predicts protein subcellular localization,⁹ (7) Secreted Protein Database or SPD (<http://spd.cbi.pku.edu.cn/>), which contains information on secreted proteins from human, mouse, and rat proteomes, including sequences from SwissProt, Trembl, Ensembl, and RefSeq,¹⁰ and (8) Signal Peptide Database (<http://www.signalpeptide.de/>), which provides signal peptide sequences for mammals containing more than 2000 confirmed sequences. From these tools, we were able to identify proteins that represent the most likely candidates for the secretome. In addition to the above current bioinformatics tools, we have implemented a publicly available Gene Ontology (GO) web-based tool (<http://ebdrup.biosustain.dtu.dk/gocho>) for annotating gene products from our data set.

CHO cells are widely used for the production of monoclonal antibodies (mAbs) and other heterologous proteins. While these mAbs can be purified, the purified proteins are often accompanied by a number of additional CHO host cell proteins (HCPs). These HCPs represent contaminants in the product mAb and must be removed during one of the purification steps. However, not all CHO host cell proteins may be removed from the end product. Unfortunately, these HCP impurities⁴ can have immunogenic effects¹¹ and also can affect product quality and stability. They can cause the formation of undesired product variants due to enzymatic activities of some HCP species such as protease and disulfide reductase.^{4a,12} For this reason, it is very important to characterize the HCPs and if possible develop methodologies for eliminating them from the product mix.^{4a} The targeted removal of HCPs from CHO cell cultures will require greater knowledge of the proteins' identity and characteristics and also represents a regulatory requirement.¹³ The current methods for quantitative estimation of HCPs lack detailed information regarding the properties or composition of the HCPs from CHO cells.¹⁴

The bioinformatics strategies presented in this study in addition with the developed CHO GO tool will help to better identify and characterize known and possible unknown CHO HCPs. In addition, this study will serve as basis for understanding the CHO secretory machinery since it categorizes the proteins containing N-terminal signal peptides and

transmembrane domains for compartmentalization as well as GO ontology information for translocation, protein folding, O-glycosylation, and N-glycosylation in the endoplasmic reticulum. A more complete understanding of the CHO secretome will facilitate current bioprocessing methodologies and provide insights into how to enhance secretory processes from CHO cells in the future.

2. MATERIALS AND METHODS

2.1. CHO Cell Samples and Isolation Materials

The CHO-K1 (CCL-61) cell line was obtained from ATCC (Manassas, VA). F-12K medium, fetal bovine serum (FBS), L-glutamine, nonessential amino acids, and DPBS were obtained from GIBCO (Grand Island, NY). Sequencing grade trypsin enzyme was purchased from Promega (Madison, WI) and the BCA protein assay kit from Thermo Scientific Pierce (Rockford, IL). Other reagents used were Tris(2-carboxyethyl) phosphine (TCEP) (Pierce, Rockford, IL), and ultrafilters (Waters, Milford, MA). All the other chemicals used in this study were purchased from Sigma-Aldrich (St. Louis, MO).

2.2. Cell Culture and Protein Lysate Preparation

CHO-K1 was cultured in supplemented 10% FBS, 1% nonessential amino acids, 2 mM L-glutamine, and F-12K media in a 37 °C incubator under 5% CO₂. After 80% confluency was reached (around a total of 18 million cells), the media was decanted, and cells were washed six times with 15 mL of PBS. Subsequently, the cells were starved for 12 h with serum-free media. The supernatant was collected after 12 h, with more than 96% cell viability, and the proteins were concentrated by centrifugation with 3 kDa ultrafilters.

2.3. In-Solution Digestion

A BCA assay was used to determine the protein concentration of both supernatant (550 µg/plate) and whole-cell lysates (5 mg/plate). Filter aided sample preparation (FASP) method¹⁵ was used prior to digesting the proteins with trypsin enzyme (1:50 ratio) at 37 °C overnight. The digested samples were separated into 96 fractions with a bRPLC method adapted from Wang et al.¹⁶ The 96 fractions were collected and concatenated into 12 fractions by merging the samples.³ⁿ The experiment was replicated with two CHO cell cultures.

2.4. LC–MS/MS Analysis

To analyze various fractions from CHO cell protein digests, 12 different LC–MS/MS analyses were performed on an LTQ-Orbitrap Velos (Thermo Electron, Bremen, Germany) mass spectrometer with an attachment of Eksigent 2D nanoflow LC system. The samples were dissolved in 8 µL of solvent, and 7.5 µL of it was used for injection. The reverse phase LC system used consisted of two parts: a peptide trap column (75 µm × 2 cm) and an analytical column (75 µm × 10 cm), which were both packed with Magic AQ C18 material (5 µm, 120 Å, www.michrom.com). After elution, the peptides were sprayed directly into an LTQ Orbitrap Velos at 2.0 kV, with a flow rate of 300 nL/min, using an electrospray (internal diameter: 8 µm) emitter tip (New Objective, Woburn, MA) with a capillary temperature of 200 °C. The entire tandem-MS analysis was carried out in Orbitrap instrument at 60 000 and 7500 resolution (measured at *m/z* 400) for precursor and the

fragment ions, respectively. FTMS full MS and MS_n AGC target were set to 1 million and 50 000 ions, respectively. Additionally, survey scans were acquired from m/z ratio of 350–1800 with up to 15 peptide masses (precursor ions) individually isolated with a 1.9 Da window and fragmented (MS/MS) using a collision energy of 35% in a higher collision dissociation (HCD) cell and 30 s dynamic exclusion. Minimum signal requirement for triggering an MS2 scan was set to 2000, and the first mass value was fixed at m/z ratio of 140. An ambient air lock mass was set at m/z ratio of 371.10123 for real-time calibration.¹⁷ Monoisotopic precursor mass selection and rejection of singly charged ion criteria were enabled for the MS/MS analysis. FTMS and FTMS/MS resolutions were set at 60 000 and 7500 at 400 m/z , respectively.

2.5. Database Searching and MS/MS Data Analysis

To analyze the MS/MS data, Mascot search engine and RefSeq annotation of CHO cells were used from the CHO genomic sequence.¹⁸ For false discovery rate (FDR) calculations, target Decoy PSM Validator node was used. In the decoy database search, a strict target FDR was set at 1% and relaxed was set at 5%. In the search engine, semitryptic enzyme specificity allowing maximum two missed cleavages, with precursor ions required to fall within 15 ppm of projected m/z values, and the mass tolerance for fragments ions was 0.03 Da was chosen. The variable modifications included oxidation (M + 15.996), deamidation (NQ), phospho (ST), phospho (Y), and pyroglut-amine (N-terminal Q, 17.027). Moreover, a fixed modification of carbamido-methylation (C + 57.021) was identified. MS raw files were charge deconvoluted and processed for deisotoping using Xtract and MS2-processor spectrum processor in addition to default spectrum selector node in Proteome Discoverer.

2.6. GO Annotation

For finding GO annotation of the secreted proteins, GO Cross Homology was obtained using GO-CHO. The GO-CHO platform takes a list of full length gene names and searches for GO terms in related organisms. In this project, we used mouse, human, and rat GO annotation.

GO-CHO was built using the Django web framework (<https://www.djangoproject.com/>), and it uses up-to-date GO annotation from <http://geneontology.org/>.¹⁹ It is freely available at <http://ebdrup.biosustain.dtu.dk/gocho/>.

2.7. Subcellular Localization and Protein Sequence Analysis

For determining subcellular localization of the identified protein sequences, we implemented a coupled use of the amino-acid sequence-based predictors, TargetP, SignalP, SecretomeP, TMHMM, Phobius, and WoLF PSORT, to increase our confidence in classifying secreted proteins.²⁰ Default D-cutoff values were chosen to optimize the performance of the search in SignalP. To increase specificity, default cutoff was used in TargetP. Normal prediction method was used in Phobius to predict subcellular localization of the proteins. Along with these predictors, an open access Secreted Proteins Database¹⁰ was also used to find out the secreted proteins from other eukaryotes. Additionally, mammalian signal peptides were obtained from an online database: Signal peptide Web site (<http://www.signalpeptide.com/>).

2.8. GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Analyses

GO terms and corresponding genes were found as described previously. KEGG pathways and corresponding genes were downloaded from KEGG Web site (<http://www.genome.jp/kegg/>). Programming tasks were performed using MATLAB version 2010a (Natick, Massachusetts: The MathWorks Inc., 2010). Enrichment *P*-values outcome is essentially a hyper-geometric distribution calculated using MATLAB's `hygecdf` and `hygepdf` functions.

2.9. Immunogenicity Prediction

The publicly available tool (<http://tools.immuneepitope.org/immunogenicity/>) was used for predicting immunogenicity of the proteins based on peptide sequences.

3. RESULTS

To characterize the CHO supernatant proteome, proteins were isolated from the cell culture broth, analyzed by MS, and subjected to multiple bioinformatics analyses as outlined in Figure 1. During the bioinformatics analysis, a number of filters were implemented including SignalP, TargetP, SecretomeP, WoLF PSORT, TMHMM, and Phobius along with databases such as SPD and Signal peptide database to identify proteins that are present in the plasma membrane and secreted into the extracellular environment of CHO cells.

The filtering steps will be described in detail in subsequent sections. Subsequently, we did GO annotation to categorize proteins to remove exclusively intracellular proteins and further select those that are functionally defined to be in the extracellular environment. In addition to functionally categorizing the CHO supernatant, relative quantification of the proteins based on a variant of spectral counting method—normalized spectral abundance factor (NSAF) values was done to identify and characterize high-abundance proteins, which could be potential HCPs. Spectral counting method encompasses several variants such as NSAF,²¹ distributed normalized spectral abundance (dNSAF),²² normalized spectral index (SIN),²³ and exponentially modified protein abundance index (emPAI).²⁴ In this study, we used NSAF values for quantifying the proteins based on the demonstrated capability of NSAF method being highly reproducible.²⁵ Furthermore, novel host proteins from this study were examined, and their immunogenicity prediction was made to identify novel immunogenic CHO proteins. Each of these steps and results of this analysis are described in greater detail in the following sections.

3.1. CHO Superome (CHO-SO) Protein Extraction, MS Experiment, and Data Analysis

Common approaches for identifying secreted proteins involve proteomic analysis of conditioned culture medium from the cell type of interest. In one approach, cells are grown in serum-bearing medium. However, this method usually necessitates extensive fractionation of proteins/peptides to detect low-abundance secreted proteins among thousands of high-abundance serum proteins.²⁶ An alternative to this approach, used in this study, was to deplete the serum after growing the cells, thereby reducing analytical interference significantly and also increasing the ability to detect relatively low-abundance secreted proteins.²⁷

To characterize the supernatant of adherent CHO-K1 cell culture, the adherent cells were grown in duplicates in serum-bearing media. After 2 days of growth, the serum was depleted, and the supernatant was collected 12 h later from the cells having more than 96% viability. The cell starvation technique is used extensively for the analysis of the secretome and secretory machinery.^{27,28} This strategy has been applied in this study to prevent FBS from masking the CHO proteins in the MS. Indeed, the results showed that albumin peptides detected in this study (FKDLGEQHFK; LSQKFPK; DLGEQHFK) are only from CHO cells and not from FBS. While some intracellular proteins may accumulate in the supernatant due to starvation or apoptosis and cell bursting, starvation is unlikely to alter the secretory machinery and superome significantly,²⁸ which was the goal of this study. As shown in Figure 1, the collected CHO supernatant was concentrated by vacuum centrifuge and ultrafiltration prior to trypsin digestion. Two-dimensional LC was used to fractionate the proteins prior to LC-MS/MS. A total of 24 fractions from the two replicates of CHO supernatant were analyzed in the mass spectrometer, and the data were analyzed using the Mascot search engine using the CHO genome for peptide and protein identification.¹⁸

To ensure high-quality data, different filtering strategies were applied: (A) cutoff of 1% FDR with more than two peptides and more than six peptide spectrum matches (PSMs) (column A, Table 1), (B) a stringent 1% FDR cutoff with one peptide and less than six PSMs (column B, Table 1), (C) cutoff of 5% FDR with more than two peptides and more than six PSMs (column D, Table 1), and (D) cutoff of 5% FDR with one peptide and less than six PSMs (column E, Table 1). Proteins meeting the above four criteria were then combined, and duplicate entries were removed. All other proteins that did not meet these four filter criteria were not used in further analysis. This resulted in 3281 CHO proteins being identified from the superome based on the combined criteria.³ⁿ Column F provides a number of unique peptides in each category, and columns G and H provide a number of PSMs in each category. A summary of the filtering results is tabulated in Table 1, and a complete overview of all the proteins is provided in the Supporting Information (complete overview of all the proteins). Of the total 3281 grouped proteins identified, 2718 exhibited at or below a 1% FDR, which is, to our knowledge, the highest number of proteins reported in the supernatant of CHO cells so far.

3.2. Relative Quantification of the Proteins in the CHO Supernatant

To elucidate the high-abundance proteins in CHO supernatant and their properties, NSAF values of each protein were calculated using the method previously described.²¹ A histogram of NSAF values of all the 3281 proteins reported in section 3.1 is shown in Figure 2, panel a. The proteins in the supernatant show a wide range of expression (NSAF) values from -21.33 to -6.33, provided in Sheet S8 (NSAF values of all 3281 identified proteins). Ninety-two proteins that showed NSAF values higher than -8.94 were outside the two standard deviation range and thus considered to be high-abundance proteins. These proteins were subjected to IPA (ingenuity pathway analysis) software to evaluate the related functional networks; an example of one such network is shown in Figure 2, panel b. From the IPA (www.ingenuity.com) software analysis, it was found that proteins such as SPARC (secreted protein acidic and rich in cysteine) and CLU (clusterin), which are both related to folding of proteins, cell survival functions, binding, and cell growth, are high in abundance.

Extracellular matrix glycoprotein SPARC, secreted by many other different cells such as osteoblasts, fibroblasts, endothelial cells, and platelets,²⁹ is involved in (a) disruption of cell adhesion,³⁰ (b) changes in cell shape,³¹ (c) inactivation of cellular responses to certain growth factors such as PDGF,³² and (d) extracellular matrix synthesis, developmental processes, angiogenesis, and binding to growth factors.

CLU, a heavily glycosylated protein³³ ubiquitously present in many tissues, functions as an extracellular chaperone that prevents aggregation of nonnative proteins and maintains partially unfolded proteins in a state appropriate for subsequent refolding by other chaperones, such as HSPA8/HSC70.

Other high-abundance proteins, PpiA (NSAF: 0.01619) and PpiB (NSAF: 0.00638) in CHO-SO, are considered to be involved in acceleration of the folding of the proteins critical to protein export and secretion. PpiA has an N-terminal uncleavable hydrophobic domain and is predicted to be an N-in C-out transmembrane domain protein.³⁴

Both SPARC and CLU have been identified as difficult to remove host cell impurities and are known to exhibit strong interactions with different monoclonal antibodies.^{4b} Some other proteins, which were present in high abundance and identified as HCPs by Valente et al., are Igfbp4 (insulin-like growth factor-binding protein 4), Vim (Vimentin), Enoa (Enolase 1), Tpm1 (tropomyosin 1), and Ldha (lactate dehydrogenase A).^{4b} Overall, out of 92 high-abundance proteins from our data set, 56 proteins are known to be HCP impurities.^{4b} A detailed list of the high-abundance proteins that are potential HCPs is provided in Sheet S9 (high-abundance proteins that are potentially HCPs).

One main concern with the presence of HCPs contamination is their potential immunogenicity. Upon comparing with previously published immunogenic³⁵ and HCPs^{4b} results, we identified 12 high-abundance proteins previously not reported as CHO HCPs or immunogenic CHO proteins. To identify the T-cell epitopes of these 12 proteins, we explored their immunogenicity using a publically available tool (<http://tools.immuneepitope.org/immunogenicity/>).³⁶ Out of these 12 proteins, eight proteins were found to contain the top 20% of potentially most immunogenic peptides with predicted immunogenicity scores higher than 0.15. Table 2 provides all of the 10 immunogenic proteins results, and the detailed results are provided in Sheet S10 (immunogenicity results).

Srsf1 contains 12 epitopes within only 198 total amino acids and belongs to a class of intrinsically disordered proteins³⁷ making it an ideal candidate to be screened by T-cells detecting aberrancies and generating an immunogenic response. Removal of these HCPs that generate immunogenic response would be desirable as part of the therapeutic proteins manufacturing process.

3.3. Addressing the Subcellular Localization of the Proteins

To functionally categorize the 3281 proteins reported in section 3.1, we implemented a number of publically available bioinformatics tools including SignalP, TargetP, SecretomeP, WoLF PSORT, TMHMM, Phobius, and also searched for the proteins in SPD as shown previously in Figure 1. Each of these tools focuses on the identification of either a signal

peptide in a given protein sequence to predict whether a protein is secreted or if a protein has a transmembrane domain. This process allowed us categorize proteins into three subcategories: (1) proteins with signal peptides, (2) proteins with transmembrane domains, and (3) proteins in the extracellular domain. This categorization provided us with a second filter for the data to segregate and retain for further analysis only those proteins known to reside in the above three categories.

A summary of the results from all of the above search engines is provided in Figure 3, panel a, and detailed results are provided in Sheet S2 (detailed results from search engines used in the study for all 3281 identified proteins). Many proteins were detected in multiple categories including containing signal peptides, as secreted proteins, or containing a transmembrane domain, resulting in a large overlap of the resulting data sets. As a result, a number of the positive results identified by one bioinformatics tools were also be identified with another tool. A simple six-way Venn diagram is provided in Figure 3, panel a showing results from the different bioinformatics tools including: (1) SignalP and Signal peptide database positive results combined providing proteins with signal peptides, (2) SecretomeP positive results of secreted proteins, (3) Phobius and TMHMM positive results with proteins containing transmembrane domains, (4) TargetP positive results also showing proteins containing a signal peptide, (5) WoLF PSORT positive results showing proteins in the extracellular space, and (6) SPD positive results with secreted proteins. SignalP predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms based on a combination of several artificial neural networks,³⁸ whereas TargetP predicts the subcellular location of eukaryotic proteins based on the predicted presence of any of the N-terminal presequences: mitochondrial targeting peptide (mTP), or secretory pathway signal peptide (SP).⁸ For the sake of simplicity, all these results are provided without the overlap details of positive results from the different tools. For example, a number of the 2398 proteins identified in SecretomeP were also identified by another search tool such as SignalP and TargetP. Overall, 66 proteins were found by all analysis tools, and 621 proteins were not identified by any of the aforementioned tools.

To identify proteins containing signal peptides, one of the approaches was to use signal peptide sequences from various species from a signal peptide Web site (<http://www.signalpeptide.de/>) to perform a stand-alone BLAST (basic local alignment search tool) analysis. This analysis compared amino acids sequences from current study with the signal peptides sequences from the aforementioned Web site. The proteins with positive results, that is, significant *E*-values, from this analysis were then combined with proteins with positive results from SignalP and TargetP analyses to further improve the prediction of signal peptides containing proteins.

Subsequently, the data sets from different searching engines were grouped into categories as shown in Figure 3, panel b. First, the positive hits from SecretomeP, SPD, and WoLF PSORT tools were combined to create a secreted proteins data set (yellow circle in Figure 3b). Second, the positive results from SignalP, TargetP, and Signal Peptide database were combined to create a data set of proteins containing signal peptides (purple circle in Figure 3b). Third, the positive results from TMHMM and Phobius were combined to create a data set of proteins containing transmembrane domains (green circle in Figure 3b). In this

integrated Venn diagram, the overlaps of proteins in different categories were noted including 447 proteins that were classified in all the three categories. An additional 262 proteins were classified in at least two of the different categories.

Elimination of the 621 proteins identified by MS resulted in a list of 2660 potentially secreted proteins (detailed in Sheet S3) (detailed results from search engines used in the study for potentially secreted proteins). Secretory proteins included signal peptides transferrin, lipoproteins, immunoglobulin domain proteins (e.g., SEMA3B, SEMA3C, and SEMA3E), collagens (e.g., COL12A1, COL16A1, and COL9A2), fibronectins (FN1), and proteoglycans (e.g., HSPG2, LEPRE1, and VCAN). These proteins are sorted in the trans-Golgi network into transport vesicles that immediately move to and fuse with the plasma membrane, releasing their contents by exocytosis.

In addition, proteins such as LRP6 (low density lipoprotein receptor-related protein 6), APP (amyloid beta A4 precursor protein), and PECAM1 (platelet/endothelial cell adhesion molecule 1) associated with receptors, cell adhesion, and binding functions, respectively, were found to be in the membrane protein data set obtained using the TMHMM and Phobius TM tools. Alternatively, extracellular proteins such as TGFB1 (transforming growth factor, beta 1), SEPT8 (septin 8), and PDGFA (platelet-derived growth factor alpha polypeptide), which perform roles as growth factors, cytokinesis regulator, and cell migration, respectively, were categorized under the secreted protein data set. Furthermore, other proteins such as SRPR (signal recognition particle receptors), cytoskeleton remodeling and organization proteins (ENAH, enabled homologue), and proteins involved in cytokinesis (SEPT6, septin 6) were grouped under proteins containing signal peptide category.

Four-hundred-forty-seven proteins found in all three categories by the search tools included ADAM17 (ADAM metallopeptidase domain 17), ALCAM (activated leukocyte cell adhesion molecule), and ERP29 (endoplasmic reticulum protein 29). Another protein in this category was SDC1 (syndecan-1), which is a single-pass, integral membrane, heparin-sulfate proteoglycan, known to contain a trans-membrane domain shed under certain conditions.³⁹ This group also included 66 experimentally identified proteins picked by all of the search engines including SPARC and SERPINC1 (also known as antithrombin-III). The protein SERPINC1 contains 465 amino acids preceded by a signal peptide of 32 amino acids.⁴⁰ This class of multifunctional proteins is involved in processes such as protease inhibition, cell-matrix interaction regulation, and inhibiting cellular proliferation,⁴¹ possibly explaining the inclusion of a signal peptide possible and their location both on the transmembrane domain and in the extracellular space.

Six-hundred-twenty-one of 3281 proteins detected by LC-MS were not detected by any of the search engines and are shown in the left-hand corners of Figure 3, panels a and b. An examination of these proteins reveals the presence of nuclear and cytoplasmic proteins such as BOD1L1 (biorientation of chromosomes in cell division 1-like 1) and CCAR1 (cell division cycle and apoptosis regulator 1), which have roles in cell division. The elimination of these candidates shows the capacity of established bioinformatics methods as a helpful tool to eliminate likely intracellular proteins for a secretome. Indeed, intracellular proteins are often detected in the supernatant from cell cultures as a result of cell bursting or lysis due

to the nature of the culture process or subsequent processing steps including centrifugation.^{4a}

3.4. GO Cross-Species Homology (GO-CHO) Database

To further analyze these predicted secreted and transmembrane proteins, we combined the 2660 proteins in the aforementioned three subcategories to obtain a first-cut of the CHO secretome (Figure 1). However, for an even stricter list of potential secretome proteins, a third filter was implemented in using GO annotation and categorization of the 2660 candidates using our newly established GO-CHO Web site, as discussed further.

The genomes of human, mouse, rat, fly, and other model organisms are very well annotated in the current literature. However, CHO genes lack such a thorough annotation, which presents a large gap in understanding the functions and cellular localization of CHO genes. However, the common ancestry between the above species and CHO cells means that a substantially large percentage of the gene products exist as homologues in the different species. This knowledge can be utilized for GO annotation of CHO genes since a gene product is likely to have a similar function and characteristics across related species. On the basis of this precept, the whole CHO genome has been functionally annotated and a web-based interface to find CHO specific gene ontologies established. The GO-CHO database can be reached at <http://ebdrup.biosustain.dtu.dk/gocho/>. In the current study, the proteins were searched against the GO-CHO database, and all related terms from well annotated species were extracted.⁴²

An example output from the GO-CHO Web site is displayed in Figure 4. A user can use GO-CHO to extract GO annotation from related well annotated species. This requires three inputs: (1) a name for the data set, (2) a list of full length gene/protein names (not symbols or abbreviations), and (3) a set of species from which to extract GO annotation. Upon clicking submit, GO-CHO searches the GO database for genes with similar names, and eventually users are provided with a searchable output of GO terms for their genes.

For the current study, the candidate 2660 proteins obtained following filtering for signal peptides, secreted proteins, and membrane proteins were input into the GO-CHO Web site to obtain the GO of the proteins. Cellular component GO terms, which provide information about the cellular location, for example, endoplasmic reticulum or Golgi apparatus, were used to filter out the proteins containing only intracellular/cytosolic GO terms such as nucleus and mitochondria. This filtering step resulted in a final filtered data set of CHO-SO containing 1015 proteins provided in Sheet S4 (final filtered data set of CHO-SO).

Many CHO-SO proteins were contained in the GO terms related to extracellular space and plasma membrane such as Dkk2 (Dickkopf-related protein 2), Klkb1 (plasma kallikrein), and Stk10 (serine/threonine-protein kinase 10). However, it is known that intracellular proteins lacking signal peptides can also accumulate in the extracellular space through cell lysis and unconventional secretion such as secreting through extracellular vesicular exosomes.⁴³ Overall, 368 proteins from the final filtered data set of 1015 proteins contained cytoplasm/cytosol GO terms, while 52 proteins were found to be contained in endoplasmic reticulum, and 71 proteins were contained in the Golgi apparatus according to GO terms.

Upon investigating the source from the previous bioinformatics filter, more than 98% of these cytosolic proteins were predicted to be secreted based on SecretomeP, and only 15% were predicted as either secreted or containing transmembrane domain by any other tool, to indicate the prediction capability limitations of these tools. However, it is important to note that a number of these proteins such as ALAD (delta-aminolevulinic acid dehydratase) contain “extracellular vesicular exosome” GO terms in addition to “cytosol” and “nucleus” cellular components, which could explain an unconventional secreting route and their subsequent inclusion in the final filtered list.

3.5. GO Enrichment Analysis

To gain a better understanding of the biological roles of the 1015 functionally annotated proteins in CHO-SO, an enrichment/overrepresentation and depletion/underrepresentation analysis of GO terms was applied using a hypergeometric distribution test. For comparison with the GO-CHO annotation of the 1015 proteins, GO annotation was performed on both the whole CHO transcriptome and proteome using the aforementioned GO-CHO Web site followed by the hyper-geometric test as a background control. CHO transcriptome data were obtained from Xu et al. study,¹⁸ and proteome data was obtained from Baycin-Hizal et al. study.³ⁿ A total of 9429 integrated genes from transcriptomics and proteomics data were used as background for finding enriched GO terms in 1015 proteins from CHO-SO data by performing the hypergeometric distribution test as previously described.⁴⁴

The results from hypergeometric distribution tests are summarized in Figure 5 in terms of the top 15 enriched GO terms in different categories for different CHO samples, also provided in Sheets S5–S7 (GO analysis: molecular functions, biological processes, and cellular components, respectively). Shown in Figure 5, panels a–c are the percentages of genes corresponding to top 15 enriched GO terms in molecular function, biological process, and cellular component, respectively, for the 1015 CHO-SO proteins.

Moreover, to compare CHO-SO cellular compartmentalization (Figure 5c) to the whole cell proteome, the hyper-geometric distribution test⁴⁵ was performed on the previously published CHO-K1 intracellular proteome containing 4391 proteins with GO cellular component terms³ⁿ with CHO genome with 13984 genes with GO cellular component terms as background¹⁸ for obtaining enriched cellular component GO terms.⁴⁴ The results of this test, Figure 5, panel d, depict that most of the enriched cellular component GO terms in intracellular proteome involve cytoplasmic space related GO terms.

The process of GO enrichment on the filtered supernatant protein data helps to focus more directly on specific classes of enriched proteins. For example, the tubulin class proteins, which are associated with the plasma membrane, such as Tuba1b, Tubb4a, and Tubg1 associated with enriched “structural constituent of cytoskeleton” molecular function (Figure 5a) are enriched in the CHO secretome.⁴⁶ Another one of the highly enriched molecular functions in CHO-SO is “calcium ion binding”, which includes Tgfb1 (transforming growth factor, beta 1) protein. Tgfb1 protein together with SMAD signaling proteins is known to be involved in IgA (immunoglobulin A) secretion.⁴⁷ The CHO intracellular proteome is known to be rich in Smad1, Smad2, Smad3, Smad4, and Smad5 proteins.³ⁿ Another enriched molecular function in CHO-SO “catalytic activity” includes plasma membrane related

proteins such as ILK (integrin-linked kinase) associated with cell junction signaling, cell adhesion, and integrin activation⁴⁸ and caveolae formation.⁴⁹ Interestingly, actin filament binding was also in the top 15 enriched GO molecular functions, including proteins such as Myh9 associated with secretion.⁵⁰

Proteins such as Sec31a and Sec23b are associated with “protein transport” as well as “vesicle mediated transport”⁵¹ and were among enriched biological processes in CHO-SO (Figure 5b). Within the enriched “intracellular protein transport” category for CHO-SO were sorting nexin family proteins such as Snx1, Snx2, and Snx3, which regulate the cell surface trafficking of growth factor receptors as well as other cytoplasmic and membrane-based proteins.⁵²

Among the enriched cellular components in CHO-SO (Figure 5c) was the “extracellular space”, which included heat shock proteins such as Hspd1, Hspe1, and Hspa13 involved in protein folding process as a part of the overall protein secretion process.⁵³ Proteins associated with the extracellular vesicular exosome include Tgfb1 (transforming growth factor, beta 1), which is secreted by CHO-K1 cells and difficult to eliminate during downstream purifications. Tgfb1 can play a key role in modulation of cellular growth, maturation and differentiation, extracellular matrix formation, homeostasis, apoptosis, and angiogenesis.⁵⁴ The contamination of clinical protein products with even minor amounts of Tgfb1 can lead to significant adverse effects.⁵⁵ For example, the presence of minute amounts of contaminating Tgfb1 can exert profound immunosuppressive effects in patients administered with human therapeutic blood products such as intravenous immunoglobulin.^{55,56}

To contrast the intracellular proteome with CHO-SO, we compared these two data sets and found that 369 proteins (approximately 36% of 1015 proteins) in CHO-SO were also found in CHO intracellular proteome. Upon close comparison, it was found that many of these 369 proteins contain cellular components such as “extracellular vesicular exosome” and “plasma membrane”. Moreover, the enriched cellular component GO terms of intracellular proteomics data set³ⁿ (Figure 5d) differ significantly from that of the CHO-SO data set (Figure 5c) in that the intracellular proteome’s three most common cellular components included nucleus, cytoplasm, and mitochondria, while that of CHO-SO included plasma membrane, extracellular space, and vesicular exosomes (Figure 5c). MEA (male enhanced antigen) is an example of the protein not identified with the intracellular proteome analysis but is instead associated with the CHO supernatant. This integral membrane protein is associated with cytoskeleton organization and includes the N-terminus on the extracellular side and the C-terminus on the cytoplasmic side. Many of the secreted proteins identified only in the CHO-SO include those such as Plat (0.0004/0.00003), Col5a2 (0.0005/0.00003), Tinagl1 (0.0008/0.00004), Csf1 (0.001/0.00001), Dag1 (0.001/0.00006), Clstn1 (0.001/0.00001), Mmp9 (0.002/0.00001), and C1ra (0.003/0.00001) (fractions in parentheses show NSAF values in CHO-SO followed by the corresponding NSAF value in intracellular proteome).³ⁿ Very low abundance values in the intracellular proteome means that many of these proteins do not accumulate significantly inside the cell, which makes it difficult to identify them with the whole cell proteomics methods.

3.6. Ingenuity Pathway Analysis

To improve our understanding of the biological functions of the 1015 proteins in CHO-SO, IPA software was used (www.ingenuity.com) to provide key functional networks. Chemotaxis of cells, exocytosis, and cell spreading, as shown in Figure 6, were some of the enriched network functions found by IPA analysis.

Importantly, many of the proteins shown in the above networks associated with “exocytosis” are also associated with secretory and signaling pathways. For example, thioredoxin (TXN), which functions to catalyze disulfide bond formation, is widely distributed and actively secreted by a variety of tissues.⁵⁷ Although TXN is known to lack a signal peptide, it follows a leaderless secretory pathway, alternative to the classical ER–golgi secretion route, and is hypothesized to translocate directly through the plasma membrane.⁵⁷ Another protein, N-ethyl-maleimide-sensitive factor attachment protein, alpha (NAPA), is a member of the soluble N-ethylmaleimide-sensitive factor (NSF) attachment proteins (SNAP or soluble NSF attachment protein) involved in diverse transport events in the secretory pathway. NAPA is functionally important for protein trafficking in the secretory pathway and may act as a SNARE (SNAP receptor) for vesicle-mediated transport events.⁵⁸ NSF together with α -SNAP dissociates the SNARE complexes that promote association and fusion of cellular membranes.⁵⁹ Another example is Perforin (PFN), which is secreted to aid in the intracellular delivery of proteases for initiating apoptosis through invagination at the plasma membrane and by promoting endocytosis of vesicles to allow membrane bound molecules into the target cells.⁶⁰ ADP ribosylation factor 6 (ARF6) from the exocytosis function is believed to mediate cytoskeletal remodeling and vesicular trafficking along the secretory pathway at the plasma membrane.⁶¹ In normal rat kidney (NRK) cells, endogenous and overexpressed ARF6 localizes to the plasma membrane and may play a role in remodeling the plasma membrane for facilitating protein secretory mechanisms.⁶² Other classes of proteins that are associated with signal transduction and trafficking of vesicles are EXOC2, EXOC8, and RHOA. The exocyst complex proteins, EXOC2 and EXOC8 (exocyst complex components), are involved in transport of the proteins from golgi apparatus to plasma membrane via the docking of exocytic vesicles with fusion sites on the plasma membrane.⁶³ Alternatively, RHOA (ras homologue family member A) regulates signal transduction pathways linking plasma membrane receptors to the assembly of focal adhesions and actin stress fibers required for the apical junction formation of keratinocyte cell–cell adhesion.⁶⁴ Another protein, SNAP23 (synaptosomal-associated protein, 23 kDa), an essential component of the high-affinity receptor for the general membrane fusion machinery and regulator of transport vesicle docking and fusion,⁶⁵ is also known to be required for integrin signaling through focal adhesion turnover in CHO cells.⁶⁶

A protein related to chemotaxis of cells and cell spreading networks is RAC1 (ras related protein), which is a plasma membrane-associated small GTPase and binds to a variety of effector proteins to regulate cellular responses such as secretory processes, phagocytosis of apoptotic cells, epithelial cell polarization, and growth-factor induced formation of membrane ruffles.⁶⁷ Among other exocytosis network signal transduction proteins serving essential cellular functionalities are AXL and B4GALT1, which are also associated with chemotaxis of cells and cell spreading networks. The cell surface form of B4GALT1

functions as a recognition molecule during a variety of cell to cell and cell to matrix interactions by binding to specific oligosaccharide ligands on opposing cells or in the extracellular matrix.⁶⁸

Several previously reported HCP impurities are associated with the previously shown exocytosis and plasma membrane projections formation networks including galectin-3 (LGALS3), vimentin (Vim), annexin A1 (Anxa1), peptidyl-prolyl cis-trans isomerase A (PPIA), transforming growth factor beta-1 (Tgfb1), laminin subunit beta-1 (Lamb1), and lactadherin (Mfge8). LAMB1 is known to have increased expression with cell age and is considered to be one of the difficult-to-remove host cell impurities.^{4b}

3.7. KEGG Pathway Enrichment Analysis of CHO-SO

Translating knowledge of these proteins into overrepresented/enriched or underrepresented/depleted biochemical pathways can be particularly meaningful. For this purpose, the KEGG database,⁶⁹ which arranges genes/proteins into specific pathways, was used to map proteins to corresponding pathways. To find enriched and depleted KEGG pathways in the filtered data set, hypergeometric analysis was performed based on known transcriptomics and proteomic data sets.¹⁸ The hypergeometric analysis revealed 111 pathways including focal adhesion, tight junction, and synaptic vesicle cycle to be significantly overrepresented (p -value <0.05) in CHO-SO. Adherence of cells to the extracellular matrix through focal adhesion pathway is known to modulate many cellular functions such as cell survival, cell proliferation, and cell motility. Besides these functions, it is also known to take part in regulation of secretion of proteins by modulating cytoskeletal rearrangements and cell spreading during secretion.⁷⁰

Shown in Figure 7 is the focal adhesion pathway with genes/proteins found in CHO-SO, and the transcriptome and proteome combined data are highlighted in different colors. Genes such as GF (growth factor) and Grb2 (growth factor receptor-bound protein 2), which are growth factor receptor bound proteins, were found in the CHO-SO. Calpain, which is known to degrade proteins including myelin, has been observed to be secreted into the extracellular medium of human cells.⁷¹ Paxillin functions to adhere cells to the extracellular matrix,⁷² and VASP is associated with actin filaments and focal adhesions that form the interface between the cytoskeleton and the extracellular matrix.⁷³

The Rac (ras-related C3 botulinum toxin substrate) subfamily of proteins is known to occupy a number of cellular locations including endosomes and caveolar vesicles, while family member Rac1 is associated with lipid rafts and focal adhesion complexes.⁷⁴ Other proteins found in the CHO-SO data include MEK1 (mitogen-activated protein kinase), which is activated by Rac through PAK (p21 protein-activated kinase). MEK1 is known to be a regulator of focal adhesion dynamics and cell migration.⁷⁵ We hypothesize that activation of Rac is necessary for activation of MEK1 and might regulate adhesion dependent cell growth.⁷⁶ Myosin light chain kinase (MLCK), known to be a main regulator of permeability in tight junction, was also found to be associated with the CHO-SO data.⁷⁷

4. DISCUSSION

The present study provides the most detailed analysis to date of the supernatant and secretome emerging from CHO cells, the most important production host in biotechnology. Because of masking by high abundance cellular proteins, low-abundance proteins present in the supernatant can be difficult to identify and characterize. However, using two-dimensional LC-based MS, we were able to identify both low-abundance proteins in CHO supernatant along with the high-abundance components that are secreted or released into the medium. Moreover, in this study, we developed a computational strategy for identifying and characterizing the secreted proteins in supernatant using multiple bioinformatics tools. For the first time, independent bioinformatics approaches were implemented to evaluate the CHO supernatant data containing 3281 proteins. With this approach, SignalP, TargetP, SecretomeP, Phobius, WoLF PSORT, and TMHMM were implemented to filter the data set to obtain predicted secreted proteins.

In this study, several bioinformatics tools such as SignalP, WoLF PSORT, TargetP, Phobius, and TMHMM were combined to find out potentially secreted proteins in CHO cells. Previous studies have evaluated the prediction accuracies of these tools.^{20,78} In one such study, SignalP, TargetP, and Phobius were evaluated for determining advantage of combining the predictions from different tools based on mammalian data, and it was proposed that using TargetP along with SignalP increases prediction accuracy to more than 95%.⁷⁹ Moreover, it was found that combining TMHMM with PSORT can improve the prediction accuracy to more than 85%.⁸⁰ This combination approach in conjunction with data from SwissProt, Trembl, Ensembl and Refseq databases was the basis for construction of a secreted protein database;^{10,80} however, there has not been any report for the categorization of CHO secreted proteins. Moreover, it has been observed that SignalP/TMHMM have higher specificity than Phobius for predicting secreted proteins from experimental data.⁸¹ Overall, the individual bioinformatics tools each exhibit different strengths in accurately predicting secreted proteins, with Phobius exhibiting most accurate predictability (82.8% for animals).²⁰ However, as discussed earlier, it has been found that combining several of these tools significantly increases the prediction accuracies of secreted proteins.²⁰ In addition to these tools, GO was also extensively used to identify secreted proteins in this study. GO has been shown to accurately identify nonclassical secretory proteins, primarily because of percentage of the nonclassical secretory proteins annotated by GO being higher than that of classical secretory proteins.⁸² Use of GO along with above bioinformatics methods provides the advantage of improving the prediction capability of both classical and nonclassical secreted proteins.

It has been found that each of these tools can provide different levels of accuracies in terms of specificity. Specificity is calculated as the percentage of number of true negatives (when the prediction tool correctly predicts a known negative observation as negative) divided by sum of total number of true negatives and total number of false positives (when the prediction tool predicts positive for a known negative observation). In this study, multiple tools were used for identifying secreted proteins, and to increase the overall specificity of the predictions, the results from different tools were combined in different groups, as shown in Figure 3, panel b. As mentioned earlier, SignalP and TargetP have reported specificities,

which is a measure of true negatives, of 78.4% and 71.6%, respectively. However, combining Phobius, WoLF PSORT, TMHMM, and TargetP can serve to increase the specificity of predictions to 97.3%.²⁰

The implementation of these bioinformatics tools resulted in 2660 potentially secreted proteins. These proteins were further filtered based on a newly established GO-CHO annotation tool to find out the cellular components of these proteins and eliminate exclusively intracellular proteins, which may be released due to cell bursting or lysis. This filtered the results to a 1015 member secreted protein data set, which we termed as superome of CHO cells (CHO-SO). Moreover, by performing a search against the full gene name of a selected organism, GO-CHO can associate GO terms with genes from species other than CHO cells to facilitate filtering of the secretome.

Overall, the final filtered CHO-SO data set consists of proteins related to a number of functions including protein folding, binding, cellular growth, and remodeling by acting as growth factors, ligands, and receptors among other purposes.⁸³ For example, TGFB1, SEPT8, and PDGFA function as growth factors, cytokinesis regulators, and cell migration proteins, respectively. Membrane proteins such as LRP6, APP, and PECAM1 serve receptor, cell adhesion, and binding functions. Proteins with signal peptides such as SRPR, ENAH, and SEPT6 are involved in signal recognition, cytoskeleton remodeling and organization, and cytokinesis.

The proteins in our data set identified in the extracellular space but lacking a signal peptide may be released into the supernatant through various mechanisms frequently referred to as “unconventional protein secretion”. These secretory mechanisms may include plasma membrane blebbing with shedding of vesicles into extracellular space or may be based on molecular mechanisms involving release of membrane-bound vesicles. Indeed, GO enrichment analysis showed that one of the significantly enriched GO cellular components of CHO-SO proteins was extracellular vesicular exosomes containing proteins such as galectins, which are unconventionally secreted,⁸⁴ as well as other proteins such as fibroblast growth factor receptor and macrophage migration inhibitory factor (MIF). This group of unconventionally secreted proteins has been shown to be influenced by external molecules such as lipoglycans.^{84,85} For example, MIF secretion can be triggered by lipopolysaccharides,⁸⁶ which may also accumulate in the supernatant of CHO cells.

Furthermore, it was found that CHO-SO is enriched in GO terms related to functions such as exocytosis, phagocytosis, formation of plasma membrane projections, and cell spreading. These GO terms contain proteins such as connective tissue growth factor (CTGF) related to cell adhesion and secretion. Moreover, CHO-SO proteins related to these functions were also associated with diverse transport events in the secretory and signaling pathways including cytoskeletal remodeling and vesicular trafficking through exocytic vesicles. Indeed, exocytic membrane trafficking pathways are known to be mediated by a series of vesicular intermediates and facilitated by specific events such as targeting, docking, and fusion. It is known that some of these protein components may constitute specificity of a single transport step, but other protein components are probably common to several transport steps leading to secretion of proteins.⁸⁷

Subcellular localization bioinformatics tools indicating such factors as the presence of signal peptides are complementary to GO annotation tools that indicate cellular components in specifying the cellular compartments of supernatant proteins from CHO. For example, WoLF PSORT provides compartmentalization of proteins within the cell such as ER and golgi apparatus in addition to cytoplasm, plasma membrane, and extracellular domain. Likewise, the GO-CHO annotation tool provides the GO, which in-turn provides all the GO terms associated with a protein, which aids in determining cellular locations including ones not reported by WoLF PSORT. For example, the protein transglutaminase 2 (TGM2) is predicted to be in nucleus, cytosol, plasma membrane, peroxisomes, and mitochondria by WoLF PSORT, whereas the GO terms related to this protein include: “mitochondrion, plasma membrane, extracellular vesicular exosome, proteinaceous extracellular matrix, and cytoplasm”. Thus, TGM2 may very well accumulate in the supernatant as a result of release from an exosome. Such a complementary compartmentalization approach can, in addition to identifying the secretome, also be useful in building models of different cellular compartments for CHO cells.

Finally, the collection of extracellular proteins found in CHO cell cultures, of both high and low abundance, whether actively secreted or released following cell lysis or other mechanisms, represents potential HCP impurities if they are copurified with drug product and as such need to be identified and reduced as much as possible in the final product. These impurities can be especially problematic as they have the potential to cause antigenic effects in human patients.¹¹ In addition to intrinsically disordered proteins such as serine/arginine-rich splicing factor 1 (Srsf1), which may generate an immunogenic response, heat shock proteins (HSPs), which act as chaperones for protein synthesis and degradation, can be present in CHO-SO. It is recognized that HSPs are major immunogens in the innate response and exhibit a dual role in innate and acquired immunity. Specifically, HSPs can transport proteins into antigen presenting cells or serve as immunogens themselves, thereby enabling the initiation of specific acquired immunity against HCPs.⁸⁸ Therefore, it is crucial to identify the proteins that can be found in the supernatant and also to understand their potential to serve as impurities in drug product. Slight differences have been found in between different CHO cells in genomic analysis, which may also reflect a certain level of changes in the CHO-SO.⁸⁹ However, we believe that by using 2D LC technique, we were able to cover the protein profiling of the supernatant that might be similar in other CHO species as well.

5. CONCLUSIONS

This study combined quantitative LC/LC-MS/MS results with GO and protein sequence analyses using multiple bioinformatics tools to elucidate a data set characteristic of the secretome from CHO cells. Qualitative and quantitative information regarding the secreted proteins in CHO cell culture may aid bioprocess development of therapeutics since these proteins can affect purification processing due to their presence as contaminants, serving to proteolyze or aggregate target proteins, and in a few cases acting as potential immune stimulating agents.

Along with the identification of immunogenic proteins and using various bioinformatics tools to determine subcellular localization of supernatant proteins, we have also established a publically available web-based tool, GO-CHO, for determining GO of CHO cells. CHO-SO will provide the most comprehensive information available to date of the assumed microenvironment and subcellular compartmentalization of the CHO cells. The types of proteins observed in this study include multiple categories and mechanisms for protein secretion by both conventional secretion mechanism as well as unconventional secretion through vesicular exosomes. The proteins observed in the secretome are enriched in a number of functions including catalytic activity or growth factors involved in the focal adhesion process. Studying the roles of these proteins will provide a more in-depth understanding of the CHO cell physiology as well as better understanding of the compartmental microenvironments of CHO cells. Such studies will also enable identification of HCPs, which can represent contaminants in the cell culture purification process as well as potential immunogens. The repertoire of the proteins secreted by the CHO cells will provide a thorough overview of the secretory mechanism of CHO cells and may provide insights into mechanism to increase secretion of target proteins and to lower contamination by undesirable HCPs going forward.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for the study was provided by MedImmune LLC and the Intramural program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

References

1. Walsh G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol.* 2010; 28:917–24. [PubMed: 20829826]
2. Walsh G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol.* 2014; 32:992–1000. [PubMed: 25299917]
3. (a) Beckmann TF, Kramer O, Klausing S, Heinrich C, Thute T, Bunttemeyer H, Hoffrogge R, Noll T. Effects of high passage cultivation on CHO cells: a global analysis. *Appl Microbiol Biotechnol.* 2012; 94:659–71. [PubMed: 22331235] (b) Bonner MK, Poole DS, Xu T, Sarkeshik A, Yates JR 3rd, Skop AR. Mitotic spindle proteomics in Chinese hamster ovary cells. *PLoS One.* 2011; 6:e20489. [PubMed: 21647379] (c) Carlage T, Kshirsagar R, Zang L, Janakiraman V, Hincapie M, Lyubarskaya Y, Weiskopf A, Hancock WS. Analysis of dynamic changes in the proteome of a Bcl-XL overexpressing Chinese hamster ovary cell culture during exponential and stationary phases. *Biotechnol Prog.* 2012; 28:814–23. [PubMed: 22556165] (d) Doolan P, Meleady P, Barron N, Henry M, Gallagher R, Gammell P, Melville M, Sinacore M, McCarthy K, Leonard M, Charlebois T, Clynes M. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnol Bioeng.* 2010; 106:42–56. [PubMed: 20091739] (e) Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA. De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat Methods.* 2012; 9:1207–11. [PubMed: 23142869] (f) Feng HT, Sim LC, Wan C, Wong NS, Yang Y. Rapid characterization of protein productivity and production stability of CHO cells by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom.* 2011; 25:1407–12. [PubMed:

- 21504006] (g) Feng HT, Wong NS, Sim LC, Wati L, Ho Y, Lee MM. Rapid characterization of high/low producer CHO cells using matrix-assisted laser desorption/ionization time-of-flight. *Rapid Commun Mass Spectrom.* 2010; 24:1226–30. [PubMed: 20391592] (h) Kantardjieff A, Jacob NM, Yee JC, Epstein E, Kok YJ, Philp R, Betenbaugh M, Hu WS. Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. *J Biotechnol.* 2010; 145:143–59. [PubMed: 19770009] (i) Meleady P, Doolan P, Henry M, Barron N, Keenan J, O'Sullivan F, Clarke C, Gammell P, Melville MW, Leonard M, Clynes M. Sustained productivity in recombinant Chinese hamster ovary (CHO) cell lines: proteome analysis of the molecular basis for a process-related phenotype. *BMC Biotechnol.* 2011; 11:78. [PubMed: 21781345] (j) Meleady P, Gallagher M, Clarke C, Henry M, Sanchez N, Barron N, Clynes M. Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells. *J Biotechnol.* 2012; 160:251–62. [PubMed: 22445466] (k) Slade PG, Hajivandi M, Bartel CM, Gorfien SF. Identifying the CHO secretome using mucin-type O-linked glycosylation and click-chemistry. *J Proteome Res.* 2012; 11:6175–86. [PubMed: 23140450] (l) Wong NS, Wati L, Nissom PM, Feng HT, Lee MM, Yap MG. An investigation of intracellular glycosylation activities in CHO cells: effects of nucleotide sugar precursor feeding. *Biotechnol Bioeng.* 2010; 107:321–36. [PubMed: 20506284] (m) Hayduk EJ, Choe LH, Lee KH. A two-dimensional electrophoresis map of Chinese hamster ovary cell proteins based on fluorescence staining. *Electrophoresis.* 2004; 25:2545–56. [PubMed: 15300775] (n) Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, Sarkaria V, Kumar A, Wolozny D, Colao J, Jacobson E, Tian Y, O'Meally RN, Krag SS, Cole RN, Palsson BO, Zhang H, Betenbaugh M. Proteomic analysis of Chinese hamster ovary cells. *J Proteome Res.* 2012; 11:5265–76. [PubMed: 22971049]
4. (a) Valente KN, Schaefer AK, Kempton HR, Lenhoff AM, Lee KH. Recovery of Chinese hamster ovary host cell proteins for proteomic analysis. *Biotechnol J.* 2014; 9:87–99. [PubMed: 24039059] (b) Valente KN, Lenhoff AM, Lee KH. Expression of difficult-to-remove host cell protein impurities during extended Chinese hamster ovary cell culture and their impact on continuous bioprocessing. *Biotechnol Bioeng.* 2015; 112:1232–42. [PubMed: 25502542]
 5. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng, Des Sel.* 2004; 17:349–56. [PubMed: 15115854]
 6. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001; 305:567–80. [PubMed: 11152613]
 7. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* 2004; 338:1027–36. [PubMed: 15111065]
 8. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000; 300:1005–16. [PubMed: 10891285]
 9. (a) Nakai K, Kanehisa M. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics.* 1992; 14:897–911. [PubMed: 1478671] (b) Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007; 35:W585–7. [PubMed: 17517783]
 10. Chen Y, Zhang Y, Yin Y, Gao G, Li S, Jiang Y, Gu X, Luo J. SPD—a web-based secreted protein database. *Nucleic Acids Res.* 2004; 33:D169–73. [PubMed: 15608170]
 11. Singh SK. Impact of product-related factors on immunogenicity of biotherapeutics. *J Pharm Sci.* 2011; 100:354–87. [PubMed: 20740683]
 12. (a) Gao SX, Zhang Y, Stansberry-Perkins K, Buko A, Bai S, Nguyen V, Brader ML. Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnol Bioeng.* 2011; 108:977–82. [PubMed: 21404269] (b) Kao YH, Hewitt DP, Trexler-Schmidt M, Laird MW. Mechanism of antibody reduction in cell culture production processes. *Biotechnol Bioeng.* 2010; 107:622–32. [PubMed: 20589844] (c) Tait AS, Hogwood CE, Smales CM, Bracewell DG. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. *Biotechnol Bioeng.* 2012; 109:971–82. [PubMed: 22124969] (d) Trexler-Schmidt M, Sargis S, Chiu J, Sze-Khoo S, Mun M, Kao YH, Laird MW. Identification and prevention of antibody disulfide bond reduction during cell culture manufacturing. *Biotechnol Bioeng.* 2010; 106:452–61. [PubMed: 20178122] (e) Wang X, Hunter AK, Mozier NM. Host cell proteins in

biologics development: Identification, quantitation and risk assessment. *Biotechnol Bioeng.* 2009; 103:446–58. [PubMed: 19388135]

13. (a) Points To Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use. Vol. 20. U.S. Food and Drug Administration, Center for Biologics Evaluation and Research; Rockville, MD: 1997. p. 214-43.(b) Aboulaich N, Chung WK, Thompson JH, Larkin C, Robbins D, Zhu M. A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies. *Biotechnol Prog.* 2014; 30:1114–24. [PubMed: 25044920] (c) Levy NE, Valente KN, Choe LH, Lee KH, Lenhoff AM. Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing. *Biotechnol Bioeng.* 2014; 111:904–12. [PubMed: 24254318]
14. Jin M, Szapiel N, Zhang J, Hickey J, Ghose S. Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnol Bioeng.* 2010; 105:306–16. [PubMed: 19739084]
15. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods.* 2009; 6:359–62. [PubMed: 19377485]
16. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, Moore RJ, Klemke RL, Camp DG 2nd, Smith RD. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics.* 2011; 11:2019–26. [PubMed: 21500348]
17. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics.* 2005; 4:2010–21. [PubMed: 16249172]
18. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol.* 2011; 29:735–41. [PubMed: 21804562]
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25:25–9. [PubMed: 10802651]
20. Min XJ. Evaluation of Computational Methods for Secreted Protein Prediction in Different Eukaryotes. *J Proteomics Bioinf.* 2010; 3:143–147.
21. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A.* 2006; 103:18928–33. [PubMed: 17138671]
22. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem.* 2010; 82:2272–81. [PubMed: 20166708]
23. Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol.* 2010; 28:83–9. [PubMed: 20010810]
24. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics.* 2005; 4:1265–72. [PubMed: 15958392]
25. McIlwain S, Mathews M, Bereman MS, Rubel EW, MacCoss MJ, Noble WS. Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinf.* 2012; 13:308.
26. (a) Finoulst I, Vink P, Rovers E, Pieterse M, Pinkse M, Bos E, Verhaert P. Identification of low abundant secreted proteins and peptides from primary culture supernatants of human T-cells. *J Proteomics.* 2011; 75:23–33. [PubMed: 21515429] (b) Colzani M, Waridel P, Laurent J, Faes E, Ruegg C, Quadroni M. Metabolic labeling and protein linearization technology allow the study of proteins secreted by cultured cells in serum-containing media. *J Proteome Res.* 2009; 8:4779–88. [PubMed: 19711906]

27. (a) Yarwood SJ, Woodgett JR. Extracellular matrix composition determines the transcriptional response to epidermal growth factor receptor activation. *Proc Natl Acad Sci U S A*. 2001; 98:4472–7. [PubMed: 11287658] (b) Mbeunkui F, Fodstad O, Pannell LK. Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. *J Proteome Res*. 2006; 5:899–906. [PubMed: 16602697]
28. Stastna M, Van Eyk JE. Investigating the secretome: lessons about the cells that comprise the heart. *Circ Cardiovasc Genet*. 2012; 5:o8–o18. [PubMed: 22337932]
29. (a) Termine JD, Kleinman HK, Whitson SW, Conn KM, McGarvey ML, Martin GR. Osteonectin, a bone-specific protein linking mineral to collagen. *Cell*. 1981; 26:99–105. [PubMed: 7034958] (b) Brekken RA, Sage EH. SPARC, a matricellular protein: at the crossroads of cell-matrix. *Matrix Biol*. 2000; 19:569–80. [PubMed: 11102747]
30. Murphy-Ullrich JE, Lane TF, Pallero MA, Sage EH. SPARC mediates focal adhesion disassembly in endothelial cells through a follistatin-like region and the Ca(2+)-binding EF-hand. *J Cell Biochem*. 1995; 57:341–50. [PubMed: 7539008]
31. Sage H, Vernon RB, Funk SE, Everitt EA, Angello J. SPARC, a secreted protein associated with cellular proliferation, inhibits cell spreading in vitro and exhibits Ca²⁺-dependent binding to the extracellular matrix. *J Cell Biol*. 1989; 109:341–56. [PubMed: 2745554]
32. Raines EW, Lane TF, Iruela-Arispe ML, Ross R, Sage EH. The extracellular glycoprotein SPARC interacts with platelet-derived growth factor (PDGF)-AB and -BB and inhibits the binding of PDGF to its receptors. *Proc Natl Acad Sci U S A*. 1992; 89:1281–5. [PubMed: 1311092]
33. Aguilar-Mahecha A, Cantin C, O'Connor-McCourt M, Nantel A, Basik M. Development of reverse phase protein microarrays for the validation of clusterin, a mid-abundant blood biomarker. *Proteome Sci*. 2009; 7:15. [PubMed: 19348683]
34. Tremillon N, Morello E, Llull D, Mazmouz R, Gratadoux JJ, Guillot A, Chapot-Chartier MP, Monlezun L, Sole V, Ginisty H, Poquet I. PpiA, a surface PPIase of the cyclophilin family in *Lactococcus lactis*. *PLoS One*. 2012; 7:e33516. [PubMed: 22442694]
35. Bailey-Kellogg C, Gutierrez AH, Moise L, Terry F, Martin WD, De Groot AS. CHOPPI: a web tool for the analysis of immunogenicity risk from host cell proteins in CHO-based protein production. *Biotechnol Bioeng*. 2014; 111:2170–82. [PubMed: 24888712]
36. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013; 9:e1003266. [PubMed: 24204222]
37. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res*. 2006; 34:305–12. [PubMed: 16407336]
38. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng, Des Sel*. 1997; 10:1–6.
39. (a) Stepp MA, Liu Y, Pal-Ghosh S, Jurjus RA, Tadvalkar G, Sekaran A, Losicco K, Jiang L, Larsen M, Li L, Yuspa SH. Reduced migration, altered matrix and enhanced TGFβ1 signaling are signatures of mouse keratinocytes lacking Sdc1. *J Cell Sci*. 2007; 120:2851–63. [PubMed: 17666434] (b) Su G, Blaine SA, Qiao D, Friedl A. Membrane type 1 matrix metalloproteinase-mediated stromal syndecan-1 shedding stimulates breast carcinoma cell proliferation. *Cancer Res*. 2008; 68:9558–65. [PubMed: 19010933]
40. Zettlmeissl G, Conradt HS, Nimtz M, Karges HE. Characterization of recombinant human antithrombin III synthesized in Chinese hamster ovary cells. *J Biol Chem*. 1989; 264:21153–9. [PubMed: 2592368]
41. Wollaston-Hayden EE, Harris RB, Liu B, Bridger R, Xu Y, Wells L. Global O-GlcNAc Levels Modulate Transcription of the Adipocyte Secretome during Chronic Insulin Resistance. *Front Endocrinol (Lausanne, Switz)*. 2015; 5:223.
42. Sanner MF. Python: a programming language for software integration and development. *J Mol Graph Model*. 1999; 17:57–61. [PubMed: 10660911]
43. Villarreal L, Mendez O, Salvans C, Gregori J, Baselga J, Villanueva J. Unconventional secretion is a major contributor of cancer cell line secretomes. *Mol Cell Proteomics*. 2013; 12:1046–60. [PubMed: 23268930]

44. Olivares-Hernandez R, Bordel S, Nielsen J. Codon usage variability determines the correlation between proteome and transcriptome fold changes. *BMC Syst Biol.* 2011; 5:33. [PubMed: 21352515]
45. Rivals I, Personnaz L, Taing L, Potier MC. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics.* 2007; 23:401–7. [PubMed: 17182697]
46. Moritz M, Agard DA. Gamma-tubulin complexes and microtubule nucleation. *Curr Opin Struct Biol.* 2001; 11:174–81. [PubMed: 11297925]
47. Li XZ, Feng JT, Hu CP, Chen ZQ. Does Arkadia contribute to TGF-beta1-induced IgA expression through up-regulation of Smad signaling in IgA nephropathy? *Int Urol Nephrol.* 2010; 42:719–22. [PubMed: 19941070]
48. (a) Honda S, Shirotani-Ikejima H, Tadokoro S, Tomiyama Y, Miyata T. The integrin-linked kinase-PINCH-parvin complex supports integrin alphaIIb beta3 activation. *PLoS One.* 2013; 8:e85498. [PubMed: 24376884] (b) Friedrich EB, Liu E, Sinha S, Cook S, Milstone DS, MacRae CA, Mariotti M, Kuhlencordt PJ, Force T, Rosenzweig A, St-Arnaud R, Dedhar S, Gerszten RE. Integrin-linked kinase regulates endothelial cell survival and vascular development. *Mol Cell Biol.* 2004; 24:8134–44. [PubMed: 15340074] (c) Tucker KL, Sage T, Stevens JM, Jordan PA, Jones S, Barrett NE, St-Arnaud R, Frampton J, Dedhar S, Gibbins JM. A dual role for integrin-linked kinase in platelets: regulating integrin function and alpha-granule secretion. *Blood.* 2008; 112:4523–31. [PubMed: 18772455]
49. Wickstrom SA, Lange A, Hess MW, Polleux J, Spatz JP, Kruger M, Pfaller K, Lambacher A, Bloch W, Mann M, Huber LA, Fassler R. Integrin-linked kinase controls microtubule dynamics required for plasma membrane targeting of caveolae. *Dev Cell.* 2010; 19:574–88. [PubMed: 20951348]
50. Nurden A, Nurden P. Advances in our understanding of the molecular basis of disorders of platelet function. *J Thromb Haemostasis.* 2011; 9:76–91. [PubMed: 21781244]
51. Renston RH, Jones AL, Christiansen WD, Hradek GT, Underdown BJ. Evidence for a vesicular transport mechanism in hepatocytes for biliary secretion of immunoglobulin A. *Science.* 1980; 208:1276–8. [PubMed: 7375938]
52. Worby CA, Dixon JE. Sorting out the cellular functions of sorting nexins. *Nat Rev Mol Cell Biol.* 2002; 3:919–31. [PubMed: 12461558]
53. Bie AS, Palmfeldt J, Hansen J, Christensen R, Gregersen N, Corydon TJ, Bross P. A cell model to study different degrees of Hsp60 deficiency in HEK293 cells. *Cell Stress Chaperones.* 2011; 16:633–40. [PubMed: 21717087]
54. (a) Ikushima H, Miyazono K. Biology of transforming growth factor-beta signaling. *Curr Pharm Biotechnol.* 2011; 12:2099–107. [PubMed: 21619537] (b) Heldin CH, Landstrom M, Moustakas A. Mechanism of TGF-beta signaling to growth arrest, apoptosis, and epithelial-mesenchymal transition. *Curr Opin Cell Biol.* 2009; 21:166–76. [PubMed: 19237272]
55. Kekow J, Reinhold D, Pap T, Ansorge S. Intravenous immunoglobulins and transforming growth factor beta. *Lancet.* 1998; 351:184–5. [PubMed: 9449877]
56. Reinhold D, Perlov E, Schrecke K, Kekow J, Brune T, Sailer M. Increased blood plasma concentrations of TGF-beta isoforms after treatment with intravenous immunoglobulins (i.v.IG) in patients with multiple sclerosis. *J Neuroimmunol.* 2004; 152:191–4. [PubMed: 15223252]
57. Rubartelli A, Bajetto A, Allavena G, Wollman E, Sitia R. Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J Biol Chem.* 1992; 267:24161–4. [PubMed: 1332947]
58. Xu Y, Wong SH, Tang BL, Subramaniam VN, Zhang T, Hong W. A 29-kilodalton Golgi soluble N-ethylmaleimide-sensitive factor attachment protein receptor (Vti1-rp2) implicated in protein trafficking in the secretory pathway. *J Biol Chem.* 1998; 273:21783–9. [PubMed: 9705316]
59. Dalal S, Rosser MF, Cyr DM, Hanson PI. Distinct roles for the AAA ATPases NSF and p97 in the secretory pathway. *Mol Biol Cell.* 2003; 15:637–48. [PubMed: 14617820]
60. Praper T, Sonnen AF, Kladnik A, Andrighetti AO, Viero G, Morris KJ, Volpi E, Lunelli L, Dalla Serra M, Froelich CJ, Gilbert RJ, Anderluh G. Perforin activity at membranes leads to invaginations and vesicle formation. *Proc Natl Acad Sci U S A.* 2011; 108:21016–21. [PubMed: 22173634]

61. (a) Yang CZ, Mueckler M. ADP-ribosylation factor 6 (ARF6) defines two insulin-regulated secretory pathways in adipocytes. *J Biol Chem.* 1999; 274:25297–300. [PubMed: 10464254] (b) Galas MC, Helms JB, Vitale N, Thierse D, Aunis D, Bader MF. Regulated exocytosis in chromaffin cells. A potential role for a secretory granule-associated ARF6 protein. *J Biol Chem.* 1997; 272:2788–93. [PubMed: 9006918] (c) Caumont AS, Galas MC, Vitale N, Aunis D, Bader MF. Regulated exocytosis in chromaffin cells. Translocation of ARF6 stimulates a plasma membrane-associated phospholipase D. *J Biol Chem.* 1998; 273:1373–9. [PubMed: 9430671]
62. Song J, Khachikian Z, Radhakrishna H, Donaldson JG. Localization of endogenous ARF6 to sites of cortical actin rearrangement and involvement of ARF6 in cell spreading. *J Cell Sci.* 1998; 111:2257–67. [PubMed: 9664047]
63. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe K, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, Hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase TO, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakebe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Yamazaki M, Watanabe K, Kumagai A, Itakura S, Fukuzumi Y, Fujimori Y, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senba T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Watanabe M, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Nakamura Y, Ohara O, Isogai T, Sugano S. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet.* 2004; 36:40–5. [PubMed: 14702039]
64. Moscow JA, He R, Gudas JM, Cowan KH. Utilization of multiple polyadenylation signals in the human RHOA protooncogene. *Gene.* 1994; 144:229–36. [PubMed: 8039707]
65. Mollinedo F, Lazo PA. Identification of two isoforms of the vesicle-membrane fusion protein SNAP-23 in human neutrophils and HL-60 cells. *Biochem Biophys Res Commun.* 1997; 231:808–12. [PubMed: 9070898]
66. Skalski M, Sharma N, Williams K, Kruspe A, Coppolino MG. SNARE-mediated membrane traffic is required for focal adhesion kinase signaling and Src-regulated focal adhesion turnover. *Biochim Biophys Acta, Mol Cell Res.* 2011; 1813:148–58.
67. Chaturvedi LS, Marsh HM, Shang X, Zheng Y, Basson MD. Repetitive deformation activates focal adhesion kinase and ERK mitogenic signals in human Caco-2 intestinal epithelial cells through Src and Rac1. *J Biol Chem.* 2007; 282:14–28. [PubMed: 17088251]
68. Amado M, Almeida R, Schwientek T, Clausen H. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim Biophys Acta, Gen Subj.* 1999; 1473:35–53.
69. (a) Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30. [PubMed: 10592173] (b) Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014; 42:D199–205. [PubMed: 24214961]
70. Kawasugi K, French PW, Penny R, Ludowyke RI. Focal adhesion formation is associated with secretion of allergic mediators. *Cell Motil Cytoskeleton.* 1995; 31:215–24. [PubMed: 7585991]
71. Deshpande RV, Goust JM, Hogan EL, Banik NL. Calpain secreted by activated human lymphoid cells degrades myelin. *J Neurosci Res.* 1995; 42:259–65. [PubMed: 8568927]
72. Schaller MD. Paxillin: a focal adhesion-associated adaptor protein. *Oncogene.* 2001; 20:6459–72. [PubMed: 11607845]

73. Aszodi A, Pfeifer A, Ahmad M, Glauner M, Zhou XH, Ny L, Andersson KE, Kehrel B, Offermanns S, Fassler R. The vasodilator-stimulated phosphoprotein (VASP) is involved in cGMP-and cAMP-mediated inhibition of agonist-induced platelet aggregation, but is dispensable for smooth muscle function. *EMBO J.* 1999; 18:37–48. [PubMed: 9878048]
74. Bustelo XR, Sauzeau V, Berenjeno IM. GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *BioEssays.* 2007; 29:356–70. [PubMed: 17373658]
75. Slack-Davis JK, Eblen ST, Zecevic M, Boerner SA, Tarcsfalvi A, Diaz HB, Marshall MS, Weber MJ, Parsons JT, Catling AD. PAK1 phosphorylation of MEK1 regulates fibronectin-stimulated MAPK activation. *J Cell Biol.* 2003; 162:281–91. [PubMed: 12876277]
76. Assoian RK, Schwartz MA. Coordinate signaling by integrins and receptor tyrosine kinases in the regulation of G1 phase cell-cycle progression. *Curr Opin Genet Dev.* 2001; 11:48–53. [PubMed: 11163150]
77. Cunningham KE, Turner JR. Myosin light chain kinase: pulling the strings of epithelial tight junction function. *Ann N Y Acad Sci.* 2012; 1258:34–42. [PubMed: 22731713]
78. Menne KM, Hermjakob H, Apweiler R. A comparison of signal sequence prediction methods using a test set of signal peptides. *Bioinformatics.* 2000; 16:741–2. [PubMed: 11099261]
79. Klee EW, Ellis LB. Evaluating eukaryotic secreted protein prediction. *BMC Bioinf.* 2005; 6:256.
80. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome.* 2003; 14:859–65. [PubMed: 14724739]
81. Tsang A, Butler G, Powlowski J, Panisko EA, Baker SE. Analytical and computational approaches to define the *Aspergillus niger* secretome. *Fungal Genet Biol.* 2009; 46:S153–S160. [PubMed: 19618504]
82. Huang WL. Ranking Gene Ontology terms for predicting non-classical secretory proteins in eukaryotes and prokaryotes. *J Theor Biol.* 2012; 312:105–13. [PubMed: 22967952]
83. Ahram M, Strittmatter EF, Monroe ME, Adkins JN, Hunter JC, Miller JH, Springer DL. Identification of shed proteins from Chinese hamster ovary cells: application of statistical confidence using human and mouse protein databases. *Proteomics.* 2005; 5:1815–26. [PubMed: 15815987]
84. Nickel W, Rabouille C. Mechanisms of regulated unconventional protein secretion. *Nat Rev Mol Cell Biol.* 2009; 10:148–55. [PubMed: 19122676]
85. Hughes RC. Secretion of the galectin family of mammalian carbohydrate-binding proteins. *Biochim Biophys Acta, Gen Subj.* 1999; 1473:172–85.
86. (a) Cleves AE, Cooper DN, Barondes SH, Kelly RB. A new pathway for protein export in *Saccharomyces cerevisiae*. *J Cell Biol.* 1996; 133:1017–26. [PubMed: 8655575] (b) Lutomski D, Fouillit M, Bourin P, Mellottee D, Denize N, Pontet M, Bladier D, Caron M, Joubert-Caron R. Externalization and binding of galectin-1 on cell surface of K562 cells upon erythroid differentiation. *Glycobiology.* 1997; 7:1193–9. [PubMed: 9455920]
87. Jackson A, Friedman S, Zhan X, Engleka KA, Forough R, Maciag T. Heat shock induces the release of fibroblast growth factor 1 from NIH 3T3 cells. *Proc Natl Acad Sci U S A.* 1992; 89:10691–5. [PubMed: 1279690]
88. Colaco CA, Bailey CR, Walker KB, Keeble J. Heat shock proteins: stimulators of innate and acquired immunity. *BioMed Res Int.* 2013; 2013:461230. [PubMed: 23762847]
89. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili I, Xu X, Wang J, Palsson BO. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat Biotechnol.* 2013; 31:759–65. [PubMed: 23873082]

Figure 1.

Overview of the process of obtaining functionally categorized CHO-SO through various filtering strategies and analysis techniques along with process of obtaining high abundance immunogenic CHO proteins.

Figure 2.

Analysis of all CHO-SO genes (a) A histogram of number of proteins for each NSAF bin with an overlaid normal distribution curve. Two standard deviations (2.4689) from the mean (-13.8317) correspond to right-hand bound of 95% confidence interval with a Log2 NSAF value of -8.8939. (b) Ninety-two proteins with Log2 NSAF values greater than -8.8939 were used to build the pathway network in IPA. The figure shows network of genes corresponding to functions—cell survival and folding of proteins.

Figure 3.

Results from bioinformatics analyses. (a) Number of positive hits as identified by different bioinformatics tools, which may also be a positive hit in another tool. (b) Proteins groups after combining positive results from SecretomeP, WoLF PSORT, and SPD as “Secreted Proteins”; positive results from SignalP, TargetP, and BLAST analysis on signal peptides database were combined as “Proteins with Signal peptides”; and positive results from TMHMM and Phobius were combined as “Membrane Proteins”.

GO CHO : GO Cross-species HOmology database

Welcome to GO CHO.

GO CHO can help you associate GO terms to your list of genes and help you filter/search for genes with specific combinations of GO terms. GO CHO performs a search against the full gene name of selected organisms and combine the results.

a) Click here to [create a new dataset](#)
Or choose one of these four datasets from the article

- [Set 1](#)
- [Set 2](#)
- [Set 3](#)
- [Set 4](#)

DTU **qrb** **Lasse Ebdrup Pedersen**
Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

GO CHO : GO Cross-species HOmology database

New Dataset Creation:

Fill out the form on the right to annotate a list of genes with GO terms from homologous species.

Step 1: Choose a unique name without special characters. The name will become part of the URL to access the dataset. E.g. if you choose the name "chosecretome1", then you will be able to access you dataset at this address
<http://ebdrup.biosustain.dtu.dk/gocho/dataview/chosecretome1>

Step 2: Copy/Paste a list of genes into the text field on the right. Use full gene names and one name per line. These gene names will be matched against the gene names of the species selected in step 3.

Step 3: Use the species search box below this text to select species. When you click a species, it will appear on the right side. Only species appearing on the right side will be used. These species are the ones from which GO annotation will be copied.

Search for species:

×

Species (click to select):

Step 1: Name for this dataset:

Step 2: List of full gene names. One name per line.

Step 3: Selected species (click to remove):

c)

Symbol	Name	Molecular Function	Cellular Component	Biological Process	Molecular Function	Cellular Component	Biological Process	Reference genes
I79_004826	aminopeptidase B	zinc ion binding; metalloprotease	extracellular space	proteolysis; leukotriene	GO:0008270; GO:0005575	GO:0005615	GO:0006508; GO:0006915	I79_004826, Amino

Figure 4.

An example output from the Web site used to find out GO of Chinese hamster and CHO cells. (a) The homepage of the Web site is accessible via <http://ebdrup.biosustain.dtu.dk/gocho/>. Clicking on “create a new data set” takes the user to a new page. (b) Users can input gene names (e.g., for which GO annotation identification is needed along with the species type on this page and submit the information). (c) The output shows gene symbol, gene name, and GO description as well as accession numbers for molecular function, biological processes, and cellular component categories.

Figure 5.

Results from GO hypergeometric distribution analysis. The percentages shown are the percentage of gene in each GO function as compared to the total genes in the GO functions. (a) Top 15 molecular functions enriched in the 1015 filtered proteins. (b) Top 15 biological processes enriched in this same list. (c) Top 15 cellular components enriched. (d) Top 15 cellular components enriched in the CHO whole cell proteome.³ⁿ

Figure 6. Functional network categories overrepresented by the extracellular space and extracellular vesicular exosome proteins. Proteins related to “cell spreading”, “chemotaxis of cells”, and “exocytosis” functions are shown.

Figure 7.

KEGG's focal adhesion pathway displaying proteins in CHO-SO in green color, proteins in CHO supernatant but not in CHO-SO in orange color, proteins in CHO transcriptome-proteome data (but not in either CHO supernatant proteins or CHO-SO) in magenta color, and genes not found in any of the above data sets in yellow color.

Table 1

Results Summary from Proteins Identified from the CHO Supernatant. The 5% FDR Row Corresponds to Entries Meeting 5% FDR but Are Not Present in the 1% FDR

	column A protein identification (peptides >2 and PSMs >6)	column B protein identification (peptides <2 and PSMs <6)	column C total proteins identified	column D unique peptides (peptides >2 and PSMs >6)	column E unique peptides (peptides <2 and PSMs <6)	column F total unique peptides	column G number of PSMs (peptides >2 and PSMs >6)	column H number of PSMs (peptides <2 and PSMs <6)	column I total number of PSMs
MS Results									
1% FDR	1340	1378	2718	13721	1904	15625	79249	9203	88452
5% FDR	8	555	563	24	587	611	56	2289	2345
total	1348	1933	3281	13745	2491	16236	79305	11492	90797

Table 2

T-Cell Epitopes Identified from High Abundance Novel CHO-SO Proteins

gene symbol	protein name	peptide	score
Srsf1	serine/arginine-rich splicing factor 1	PFAFVEFED	0.42225
Prdx6	peroxiredoxin-6	RVVFIFDPD	0.38324
Rps21	40S ribosomal protein S21	ASNRIIGAK	0.35632
Hnrnpa1	heterogeneous nuclear ribonucleoprotein A1	KRGFAFVTF	0.33462
Hist1h3a	histone H3.1	LARRIRGER	0.3343
Tpm1	tropomyosin alpha-1 chain	KQVEEELTH	0.31922
Hist1h3c	histone H3.2	QRLVREIAQ	0.31777
EIF5AL1	eukaryotic translation initiation factor 5A-1	YDCGEEILI	0.31419
Rps27a	ubiquitin-40S ribosomal protein S27a	PSDTIENVK	0.28877
Hspe1	heat shock 10 kDa protein 1	LPLFDRVLV	0.21288