Minireview
Maturation of the mammalian secretome
Jeremy C Simpson, Alvaro Mateos and Rainer Pepperkok

Address: Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse, 69117 Heidelberg, Germany.

Correspondence: Jeremy C Simpson. Email: simpson@embl.de. Alvaro Mateos. Email: mateos@embl.de

Abstract

A recent use of quantitative proteomics to determine the constituents of the endoplasmic reticulum and Golgi complex is discussed in the light of other available methodologies for cataloging the proteins associated with the mammalian secretory pathway.

The secretory pathway in mammalian cells consists of an array of membrane-bounded organelles and transport carriers through which secretory proteins move in a stepwise fashion to reach their different cellular destinations. This arrangement means that biochemical operations carried out by the pathway, such as protein folding, protein glycosylation or lipid biosynthesis, can be compartmentalized, which enables efficient and specific reactions. Because of this link between subcellular localization and function, a quantitative map of the distribution of all the protein and lipid constituents of the secretory pathway (the secretome) is an essential first step to a comprehensive molecular understanding of how the pathway functions. For the past 30 years, cell biologists and biochemists have addressed this problem by imaging immunolabeled components in intact cells and by the biochemical analysis of subcellular fractions (reviewed in [1,2]). Although these approaches have generated an enormous amount of detailed knowledge on the composition of the secretory pathway, the availability of the complete sequences of several eukaryotic genomes has recently enabled more systematic attempts to describe the secretory pathway comprehensively at the molecular level. Landmark studies in which a large, almost genome-wide, fraction of yeast open reading frames (ORFs) were tagged with green fluorescent protein (GFP) and their subcellular localizations determined in living cells have generated comprehensive localization maps for the *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* proteomes and thus also the secretory pathways in these organisms [3,4]. Related approaches in mammalian cells have also been reported [5,6], although these are still far less comprehensive than for yeast.

Subcellular fractionation followed by mass spectrometry (MS)-based analysis has also proved highly successful in mapping proteins to specific subcellular structures, such as the Golgi complex [7-9] or clathrin-coated vesicles [10,11]. Remarkably, recent advances in MS-based proteomics (reviewed in [12]) now even allow estimation of the relative abundance of proteins in a specific biochemical fraction, opening up new avenues for defining a genome-wide localization map of a mammalian proteome. Taking advantage of this technological progress, Gilchrist and colleagues [13] have recently produced an MS-based proteomic map of the major membrane-bounded entities of the mammalian secretory pathway - the endoplasmic reticulum (ER) and the Golgi complex. The significance of this well controlled piece of work is that it both complements and extends previous proteomic analyses of the mammalian secretory pathway [7-11,14].

Gilchrist *et al.* [13] used classical biochemical procedures to isolate rough ER, smooth ER and Golgi membranes from rat liver, and then assessed fraction purity by electron microscopy and enzyme activity analyses. Solubilized membranes were subjected to gel electrophoresis and quantitative tandem MS, which identified peptides that could be mapped to more than 2,000 proteins. Assignment of these proteins to 23 different functional categories allowed the *in silico* removal of 470 proteins that were probably contaminants, with almost two-thirds of these being residents of mitochondria and the plasma membrane. Throughout this study, independent samples were prepared and analyzed in triplicate, with principal coordinate analysis confirming that

Table 1

Notable experimental studies in the determination of the secretome

Year	Approach	Organelle	Number of proteins	Comment	Reference
1980	Mutagenesis	Cell-wide	23	First attempt to systematically identify secretory machinery (*S. cerevisiae*)	[27]
2000	Proteomics	Golgi	93 (588)*	First concerted effort to analyze the Golgi proteome. Cycloheximide treatment used to enrich for Golgi residents	[7]
2003	GFP-tagging	ER	296	Subcellular localization of more than 75% of the yeast proteome was determined using GFP fusions (*S. cerevisiae*)	[3]
		Golgi/TGN	118		
2004	Proteomics	Golgi	151 (421)†	Improved proteomics technologies allowed more comprehensive identification of proteins	[9]
2006	Functional assay	Cell-wide	130	First genome-wide RNAi screen to search for proteins with a functional involvement in secretion (*Drosophila* S2 cells)	[26]
2006	Proteomics	ER	229	Proteomics and protein correlation profiling of multiple organelles separated by gradient centrifugation	[14]
		ER-Golgi vesicles	220		
		Golgi	67		
2006	GFP-tagging	ER	382 (68)‡	Subcellular localization of 90% of the yeast proteome was determined using GFP fusions (*S. pombe*)	[4]
		Golgi	156 (68)‡		
2006	Proteomics	ER	832 (405)§	Most quantitative proteomic characterization of the ER, Golgi and COPI vesicles reported to date	[13]
		Golgi	193 (405)§		

*Mass spectrometry was used to identify only 93 of the 588 unique spots observed on two-dimensional gels. †From 421 identified proteins, only 151 were annotated as being either unknown or *bona fide* Golgi residents. ‡An additional 68 proteins localized to the ER and Golgi. §An additional 405 proteins were identified in both ER and Golgi fractions.

the ER and Golgi fractions were consistently distinct from one another. Further clustering of the identified proteins was facilitated by subfractionation using salt washing and Triton X-114 phase separation, which finally yielded an impressive list of 832 unique ER proteins, 193 proteins of the Golgi complex and COPI transport vesicles, and a further 405 proteins that were found in both fractions.

This seems to be the most comprehensive effort so far to elucidate the proteomes of the organelles of the mammalian secretory pathway. An impressive number of controls were incorporated at every step to give the highest degree of confidence in the lists obtained. Closer analysis of these lists reveals that the vast majority of well known residents of the organelles have been identified, for example the components of the protein folding and glycosylation machinery in the ER, and the protein-modification enzymes in the Golgi complex, in addition to more than 300 uncharacterized proteins. Of particular note is the identification of many cytoplasmic proteins that are only transiently associated with membranes, including many components of the actin and microtubule cytoskeletons. However, a significant number of likely contaminants also seem to be present in the fractions, highlighting the fact that, despite improvements in the sensitivity of MS, the limitation of this type of approach remains at the level of the organelle separation techniques. For example, the identification of proteins of the plasma membrane/endocytic machinery (such as the clathrin adaptor 2 alpha subunit (CALM) and the GTPase dynamin) in the ER fraction indicates the difficulty of separating these membranes.

How near are we to a complete secretome?

The work by Gilchrist *et al.* [13], together with the information available in the literature, raises the question of whether we are now able to define a mammalian secretome. Considering the care taken to exclude likely contaminants in this study, and comparing it with earlier proteomic analyses (Table 1), the indication is that subcellular fractionation combined with MS-based proteomics is unlikely to reveal many more new proteins that map to the organelles of the secretory pathway. Nevertheless, extrapolation of the genome-wide localization data from *S. cerevisiae* [3,15] and *S. pombe* [4] to the 30,000 ORFs of the human genome suggests that the number of human proteins associated with the ER and Golgi complex could be between 2,850 and 4,110 [16]. This is more than twice the number proposed by Gilchrist *et al.* [13], and may indicate an intrinsic lack of completeness in subcellular fractionation and MS-based approaches. One reason for this may be that during subcellular fractionation, a significant number of proteins transiently associated with the organelles under investigation may be lost to the extent that they fall below the current detection level of MS. Future improvements in MS sensitivity may help to overcome this problem. Also,

most fractionation/MS-based studies focus on a single type of tissue (predominantly brain and liver) as the material for analysis. The number of proteins associated with the ER and Golgi complex should, therefore, increase when the variety of tissues in the human body is considered.

The problem of transient protein-organelle interactions within the secretory pathway can be addressed by GFP-tagging and subcellular localization in living cells. Light-microscopy of cells expressing GFP-tagged markers provides excellent resolution and sensitivity, and can, in principle, monitor even very transient localizations lasting only seconds. Many examples of functionally significant transient interactions in the secretory pathway are known (see, for example, [17-19]). Indeed, proteins shown to interact with components of the secretory pathway in those experiments (p150^{glued} [17], γ-BAR [18], and the PICTAIRE kinases [19]) were not found by Gilchrist *et al.* [13] to be associated with any of their fractions. This shows that approaches that can also reveal transient interactions with membranes of the secretory pathway are essential if the secretome is to be completely defined.

Having defined the essential secretome components, the next step will be to go beyond basic localization studies and map these proteins to the organelles in which they function. Live-cell imaging of GFP-tagged proteins can provide sufficient spatial and temporal resolution but is, unfortunately, still limited in throughput (reviewed in [20]). Although simple cellular morphological changes can be monitored by time-lapse microscopy in a high-throughput manner [21], and sophisticated image-analysis technologies have become available to accurately determine subcellular localization (reviewed in [22]), large-scale quantitative mapping of the GFP-tagged proteins to specific organelles is still not possible, as it requires the acquisition of image data in three dimensions, which is a slow process.

Combining localization and functional studies

Functional studies may help here as they not only support the localization information but can also begin to provide information about the networks in which each protein operates. In cultured mammalian cells, protein over-expression and downregulation are the most immediate ways of studying a protein's function [23,24]. Vast and easily accessible collections of cDNAs and ORFs make over-expression possible [25], while RNA interference (RNAi) makes large-scale knockdown experiments feasible [26]. Understanding the molecular basis of the secretory pathway using overexpression and downregulation techniques has effectively been 'work in progress' for more than 25 years. The pioneering experiments were carried out in yeast [27], largely because of its genetic tractability and the fact that its genome does not have the complexity of higher eukaryotes, which have tissue-specific variation in gene expression and

extensive splice variants, for example. These first lists of candidate secretome proteins have stood the test of time, and represent much of the core secretion machinery found in all eukaryotes.

More recently, a complete genome-wide downregulation screen in *Drosophila* S2 cells was reported [26], highlighting the advent of functional screening as a means of determining the secretome in more complex organisms. Combining the information from such approaches with proteomics-based localization strategies is a potentially enormously powerful approach, as the two methods are methodologically independent yet aspire to the same goal. Indeed, of the 77 mammalian orthologs identified in the *Drosophila* screen as affecting secretion (from 130 fly candidates), a third were also identified by Gilchrist *et al.* [13]. This correspondence allows preliminary mapping of the functional effects of these proteins to a particular subcellular compartment, but the question remains as to why there is not greater overlap between these lists. Incorrect identification of orthologs across species may be one explanation, but this discrepancy is more likely to reflect the fact that functional approaches alone cannot provide comprehensive lists of the constituents of the organelles involved. Similarly, determination of localization does not directly infer function, but rather should be considered as another essential piece of information towards the goal of identifying the secretome.

The approaches outlined above are complementary to other methods that are now being applied to studying cellular composition and function on a genome-wide scale - for example, comparative proteomics and mRNA expression profiling [28]. The power of these approaches is that *in silico* data can be readily incorporated to extrapolate and predict discrete functional networks. An excellent example of this strategy is the recent definition of the 'membrome', a comprehensive listing of the key interacting components that define the membrane architecture of a specific cell type [29].

The complete secretome may still not have been identified, but the tools and technologies that will achieve this are now established and in use. As well as its intrinsic interest, the secretome is of great medical importance, because dysfunctional membrane trafficking pathways have many clinical implications [30]. Drug-discovery programs will surely become more efficient if we have already mapped all the relevant proteins to their organelle and functional interaction network.

References

1. Presley JF: **Imaging the secretory pathway: the past and future impact of live cell optical techniques.** *Biochim Biophys Acta* 2005, **1744:**259-272.
2. Bonifacino JS, Glick BS: **The mechanisms of vesicle budding and fusion.** *Cell* 2004, **116:**153-166.
3. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast.** *Nature* 2003, **425:**686-691.
4. Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, Sekido S, Kobayashi Y, Hashimoto A, Hamamoto M, Hiraoka Y, *et al.*: **ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*.** *Nat Biotech* 2006, **24:**841-847.
5. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S: **Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing.** *EMBO Rep* 2000, **1:**287-292.
6. Uhlen M, Bjorling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al.: **A human protein atlas for normal and cancer tissues based on antibody proteomics.** *Mol Cell Proteomics* 2005, **4:**1920-1932.
7. Taylor RS, Wu CC, Hays LG, Eng JK, Yates III JR, Howell KE: **Proteomics of rat liver Golgi complex: minor proteins are identified through sequential fractionation.** *Electrophoresis* 2000, **21:**3441-3459.
8. Bell AW, Ward MA, Blackstock WP, Freeman HNM, Choudhary JS, Lewis AP, Chotai D, Fazel A, Gushue JN, Paiement J, *et al.*: **Proteomics characterization of abundant Golgi membrane proteins.** *J Biol Chem* 2001, **276:**5152-5165.
9. Wu CC, MacCoss MJ, Mardones G, Finnigan C, Mogelsvang S, Yates III JR, Howell KE: **Organellar proteomics reveals Golgi arginine dimethylation.** *Mol Biol Cell* 2004, **15:**2907-2919.
10. Blondeau F, Ritter B, Allaire PD, Wasiak S, Girard M, Hussain NK, Angers A, Legendre-Guillemin V, Roy L, Boismenu D, *et al.*: **Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling.** *Proc Natl Acad Sci USA* 2004, **101:**3833-3838.
11. Borner GHH, Harbour M, Hester S, Lilley KS, Robinson MS: **Comparative proteomics of clathrin-coated vesicles.** *J Cell Biol* 2006, **175:**571-578.
12. Andersen JS, Mann M: **Organellar proteomics: turning inventories into insights.** *EMBO Rep* 2006, **7:**874-879.
13. Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, Nagaya H, Roy L, Gosline SJC, Hallett M, *et al.*: **Quantitative proteomics analysis of the secretory pathway.** *Cell* 2006, **127:**1265-1281.
14. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M: **A mammalian organelle map by protein correlation profiling.** *Cell* 2006, **125:**187-199.
15. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, *et al.*: **Subcellular localization of the yeast proteome.** *Genes Dev* 2002, **16:**707-719.
16. Simpson JC, Pepperkok R: **Localizing the proteome.** *Genome Biol* 2003, **4:**240.1-240.4.
17. Watson P, Forster R, Palmer KJ, Pepperkok R, Stephens DJ: **Coupling of ER exit to microtubules through direct interaction of COPII with dynactin.** *Nat Cell Biol* 2005, **7:**48-55.
18. Neubrand VE, Will RD, Mobius W, Poustka A, Wiemann S, Schu P, Dotti CG, Pepperkok R, Simpson JC: **Gamma-BAR, a novel AP-1-interacting protein involved in post-Golgi trafficking.** *EMBO J* 2005, **24:**1122-1133.
19. Palmer KJ, Konkel JE, Stephens DJ: **PCTAIRE protein kinases interact directly with the COPII complex and modulate secretory cargo transport.** *J Cell Sci* 2005, **118:**3839-3847.
20. Pepperkok R, Ellenberg J: **High-throughput fluorescence microscopy for systems biology.** *Nat Rev Mol Cell Biol* 2006, **7:**690-696.
21. Neumann B, Held M, Liebel U, Erfle H, Rogers P, Pepperkok R, Ellenberg J: **High-throughput RNAi screening by time-lapse imaging of live human cells.** *Nat Methods* 2006, **3:**385-390.
22. Glory E, Murphy RF: **Automated subcellular location determination and high-throughput microscopy.** *Dev Cell* 2007, **12:**7-16.
23. Starkuviene V, Liebel U, Simpson JC, Erfle H, Poustka A, Wiemann S, Pepperkok R: **High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic.** *Genome Res* 2004, **14:**1948-1956.
24. Simpson JC, Cetin C, Erfle H, Joggerst B, Liebel U, Ellenberg J, Pepperkok R: **An RNAi screening platform to identify secretion machinery in mammalian cells.** *J Biotech*, in press.
25. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, *et al.*: **Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.** *Proc Natl Acad Sci USA* 2002, **99:**16899-16903.
26. Bard F, Casano L, Mallabiabarrena A, Wallace E, Saito K, Kitayama H, Guizzunti G, Hu Y, Wendler F, DasGupta R, *et al.*: **Functional genomics reveals genes involved in protein secretion and Golgi organization.** *Nature* 2006, **439:**604-607.
27. Novick P, Field C, Schekman R: **Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway.** *Cell* 1980, **21:**205-215.
28. Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, et al.: **Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling.** *Cell* 2006, **125:**173-186.
29. Gurkan C, Lapp H, Alory C, Su AI, Hogenesch JB, Balch WE: **Large-scale profiling of Rab GTPase trafficking networks: the membrome.** *Mol Biol Cell* 2005, **16:**3847-3864.
30. Aridor M, Hannan LA: **Traffic jams II: an update of diseases of intracellular transport.** *Traffic* 2002, **3:**781-790.