

Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein–protein interactions

Georg H. H. Borner^a, Marco Y. Hein^{a,*}, Jennifer Hirst^{b,*}, James R. Edgar^b, Matthias Mann^a, and Margaret S. Robinson^b

^aDepartment of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; ^bCambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom

ABSTRACT We developed “fractionation profiling,” a method for rapid proteomic analysis of membrane vesicles and protein particles. The approach combines quantitative proteomics with subcellular fractionation to generate signature protein abundance distribution profiles. Functionally associated groups of proteins are revealed through cluster analysis. To validate the method, we first profiled >3500 proteins from HeLa cells and identified known clathrin-coated vesicle proteins with >90% accuracy. We then profiled >2400 proteins from *Drosophila* S2 cells, and we report the first comprehensive insect clathrin-coated vesicle proteome. Of importance, the cluster analysis extends to all profiled proteins and thus identifies a diverse range of known and novel cytosolic and membrane-associated protein complexes. We show that it also allows the detailed compositional characterization of complexes, including the delineation of subcomplexes and subunit stoichiometry. Our predictions are presented in an interactive database. Fractionation profiling is a universal method for defining the clathrin-coated vesicle proteome and may be adapted for the analysis of other types of vesicles and particles. In addition, it provides a versatile tool for the rapid generation of large-scale protein interaction maps.

Monitoring Editor

Adam Linstedt
Carnegie Mellon University

Received: Jul 8, 2014

Revised: Aug 6, 2014

Accepted: Aug 7, 2014

INTRODUCTION

Several thousand proteins are associated with the eukaryotic endomembrane system, yet the characterization of their functional interactions remains a mostly unresolved question (Havugimana *et al.*, 2012). Numerous large-scale studies have been carried out to determine the localization and interactome of individual proteins using fluorescence or affinity tags (Huh *et al.*, 2003; Ewing *et al.*, 2007), but these approaches are extremely labor intensive and also suffer from potential interference introduced by tagging. An alternative

strategy is to separate organelles by density gradient centrifugation and identify proteins associated with individual fractions through mass spectrometry (Dunkley *et al.*, 2004; Foster *et al.*, 2006). Such comparative proteomic approaches have yielded valuable insights into the composition of major organelles. However, transport vesicles, which are responsible for the selective exchange of contents between organelles, are not resolved by these methods.

Clathrin-coated vesicles (CCVs) mediate transport between the *trans*-Golgi network and endosomes and also facilitate plasma membrane endocytosis (Robinson, 2004; Faini *et al.*, 2013). Their function is critically important in all eukaryotes: in multicellular organisms, defects in CCV trafficking are embryonic lethal (Robinson, 2004; Borner *et al.*, 2007; Umasankar *et al.*, 2012) or associated with severe developmental phenotypes (Tarpey *et al.*, 2006; Montpetit *et al.*, 2008). Similarly, unicellular trypanosomes, the causative agents of sleeping sickness, are not viable without clathrin-mediated endocytosis (Allen *et al.*, 2003). To date, the CCV composition from only three sources has been characterized by proteomics: rat brain (Blondeau *et al.*, 2004), rat liver (Girard *et al.*, 2005), and HeLa cells (Borner *et al.*, 2006, 2012; Hirst *et al.*, 2012). Because standard biochemical fractionation procedures never yield “pure” fractions, an

This article was published online ahead of print in MBoC in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E14-07-1198>) on August 27, 2014.

*These authors contributed equally to this work.

The authors declare that they have no competing interests.

Address correspondence to: Georg H.H. Borner (borner@biochem.mpg.de).

Abbreviations used: CCV, clathrin-coated vesicle; MAD, median absolute deviation from the median; PCA, principal components analysis; SILAC, stable isotope labeling by amino acids in cell culture.

© 2014 Borner *et al.* This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®,” “The American Society for Cell Biology®,” and “Molecular Biology of the Cell®” are registered trademarks of The American Society of Cell Biology.

assignment of identified proteins as genuine CCV components or contaminants is highly ambiguous (Blondeau *et al.*, 2004; Girard *et al.*, 2005). Recently we showed how selective interference with CCV formation through gene silencing (Borner *et al.*, 2012) or drug-induced knockdowns (Hirst *et al.*, 2012) in conjunction with quantitative mass spectrometry can be used to define the contents of CCVs objectively and with high accuracy. A limitation of these perturbation-based approaches, however, is that they require gene knockdown or transgene expression and are thus restricted to cell types amenable to these techniques. There is no universal method for the proteomic analysis of CCVs and, more generally, transport vesicles. To fill this niche, we developed “fractionation profiling,” which is based exclusively on quantifying subcellular fractionation behavior. The method requires no chemical or genetic manipulation and reveals the composition of CCVs with high specificity and sensitivity. Furthermore, our data suggest that fractionation profiling will work in a wide variety of cell types. Although we initially designed the method for the analysis of transport vesicles, we also demonstrate its potential for discovering and characterizing protein complexes.

RESULTS

Development of the fractionation profiling approach

HeLa cells are still the only human cell type with a well-characterized CCV proteome (Borner *et al.*, 2012), and we hence developed and validated fractionation profiling in these cells. We previously showed that a “mixed vesicle fraction” can be prepared from HeLa cells by standard biochemical fractionation techniques (Borner *et al.*, 2006, 2012). This fraction is highly enriched in CCVs but also contains other membrane vesicles, as well as large cytosolic protein complexes. The aim of fractionation profiling is not to obtain a “pure” CCV fraction but to determine the subcellular fractionation behavior of CCVs and identify all proteins that share it. First, we established differential centrifugation conditions that allow the further division of the vesicle fraction into three subfractions of roughly equal total protein content (Figure 1A). In parallel, we prepared a reference fraction from metabolically “heavy” labeled cells (stable isotope labeling by amino acids in cell culture [SILAC] method; Ong *et al.*, 2002). This reference combines the contents of all three subfractions. Electron microscopy shows that CCVs and other structures have characteristic enrichment and depletion profiles across the subfractions (Figure 1B). Gel electrophoretic analysis (Figure 1, C and D) confirms that CCV-associated proteins share the same sharply defined profile, with a pronounced peak in fraction 1 and almost complete depletion from fraction 3. This signature profile is clearly resolved from coat components of major non-CCV transport vesicles and other particles, such as AP-3, AP-4, and ribosomes.

For proteomic analysis, an aliquot of the SILAC heavy reference fraction was pooled with each (SILAC light) subfraction. The experiment was then repeated with reversed metabolic labels. Thus six sample pairs were obtained and analyzed by quantitative mass spectrometry (Figure 1E). In total, we identified >4500 proteins (Supplemental Table S1), with 2827 proteins quantified across all six samples. For every protein, we determined the abundance in each subfraction relative to the reference. The resulting six ratios constitute a protein’s profile. As expected, coat components of CCVs show very closely matched profiles and are clearly distinguishable from non-CCV proteins (Figure 1F).

To sort proteins into groups with related profiles, we first performed principal components analysis (PCA). This allowed the representation of the variation in our six-dimensional data set as a two-dimensional scatter plot and clustered proteins based on profile

similarity (Figure 2). Known CCV proteins clustered tightly in the periphery of the plot. Hence fractionation profiling successfully segregated CCV from non-CCV proteins. Of importance, the cluster analysis extends to all other proteins in the data set; numerous known protein complexes with diverse functions also formed tight clusters throughout the plot, as discussed later.

The Predictor database

To harness the profiling data, we designed an interactive analysis tool (see *Materials and Methods*). The “Predictor” (Supplemental Table S1) searches the profiling data against query proteins and provides a classified output of proteins with similar profiles. Profile similarity calculations are based on distances in multidimensional log space. Of importance, the Predictor is not based on PCA. The purpose of the PCA plot in Figure 2 is to provide a graphical illustration of profile clustering. The reduction of a six-dimensional data set to a two-dimensional scatter plot necessarily causes a loss of discriminating power. As a result, some clusters of known protein complex subunits (colored dots) appear to overlap with unrelated proteins (gray dots). The Predictor, however, achieves much higher resolution than shown in Figure 2 by calculating distances in six-dimensional space. See *Materials and Methods* for further details.

Clathrin heavy chain (CHC) is the major CCV protein, and proteins with similar profiles are candidate CCV constituents. To test whether the Predictor can recover the known HeLa CCV proteome, we searched the database of 2827 complete profiles against CHC. The Predictor classified 93 hits as “very high confidence” CCV proteins; 84 of these are previously reported CCV proteins. Based on their functional annotation, eight of the remaining nine proteins are strong candidate novel CCV proteins, and only one is a definite false positive, suggesting a predictive accuracy of 90–98% (at the 60% sensitivity threshold; see Supplemental Table S2 for a detailed analysis of the predictions and new CCV proteins). A further 75 proteins were classified as “high-confidence” predictions, bringing the total number of predicted CCV proteins to 168. These include 112 known CCV proteins, as well as 45 strong candidate novel coat and cargo CCV proteins, with <7% definite false positives (at the 80% sensitivity threshold; Supplemental Table S2). Collectively these data suggest that fractionation profiling can identify known and novel CCV proteins with exceptional accuracy and coverage.

Application of fractionation profiling to *S2 Drosophila* cells reveals the first insect CCV proteome

Although *Drosophila* is a widely used model organism, the composition of its CCVs remains poorly characterized. To test whether fractionation profiling is transferable to other cell systems, we chose to investigate *Drosophila S2* cells. We applied fractionation profiling without further optimization (Figure 3A). Clathrin and associated proteins had profiles similar to those in HeLa cells (Figure 3B). For proteomic analysis, we prepared a reference and three subfractions from SILAC-labeled *S2* cells, repeated the preparation with reversed labeling, and analyzed the six sample pairs by mass spectrometry. In total, we identified >3000 proteins, of which 1799 were quantified across all six samples. PCA shows that subunits of known protein complexes form clusters, as expected (Figure 3C). Known fly CCV proteins, including clathrin, AP-1, and AP-2, formed a distinct cluster in the periphery of the plot. We then constructed a Predictor for the *S2* profiling data (Supplemental Table S3). A search against *Drosophila* clathrin heavy chain revealed a list of 29 candidate CCV proteins predicted with the highest level of confidence (Table 1). Remarkably, the human homologues of 27 of these are known CCV proteins; most of these proteins have not been characterized in

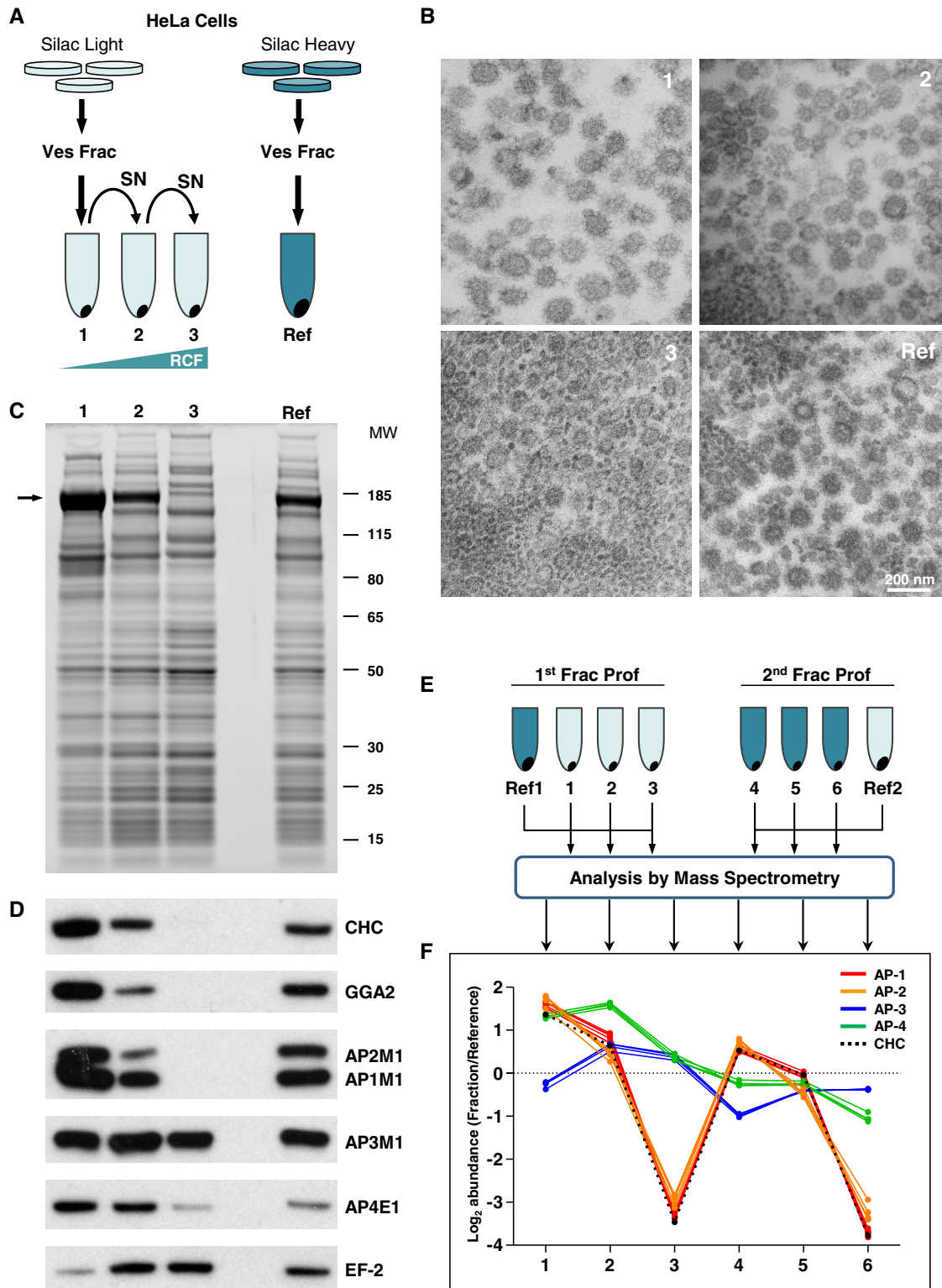


FIGURE 1: Development of the fractionation profiling approach. (A) Workflow. HeLa cells were metabolically labeled with heavy (Lys-8/Arg-10) or light (Lys-0/Arg-0) amino acid isotopes. Equivalent vesicle-enriched membrane fractions (Ves Frac) were prepared by biochemical fractionation. The light Ves Frac was subfractionated by a series of three centrifugation steps with increasing speed (pellets 1–3). In parallel, the heavy Ves Frac was centrifuged once at high speed to obtain the reference pellet. (B) Characterization of subfractions by electron microscopy. Pellets were fixed and cross-sectioned. CCVs are the predominant profiles in pellet 1 and very abundant in pellet 2. Pellet 3 contains mostly smooth membrane vesicles and nonvesicular protein particles (e.g., proteasomes and ribosomes). The reference fraction has an intermediate composition. (C) Analysis by gel electrophoresis (Coomassie stain). All fractions have distinct compositions. Clathrin heavy chain (small arrow) is enriched in fractions 1 and 2 and depleted from fraction 3. Several other prominent bands show the opposite trend. (D) Analysis by Western blot. Clathrin heavy chain (CHC) and

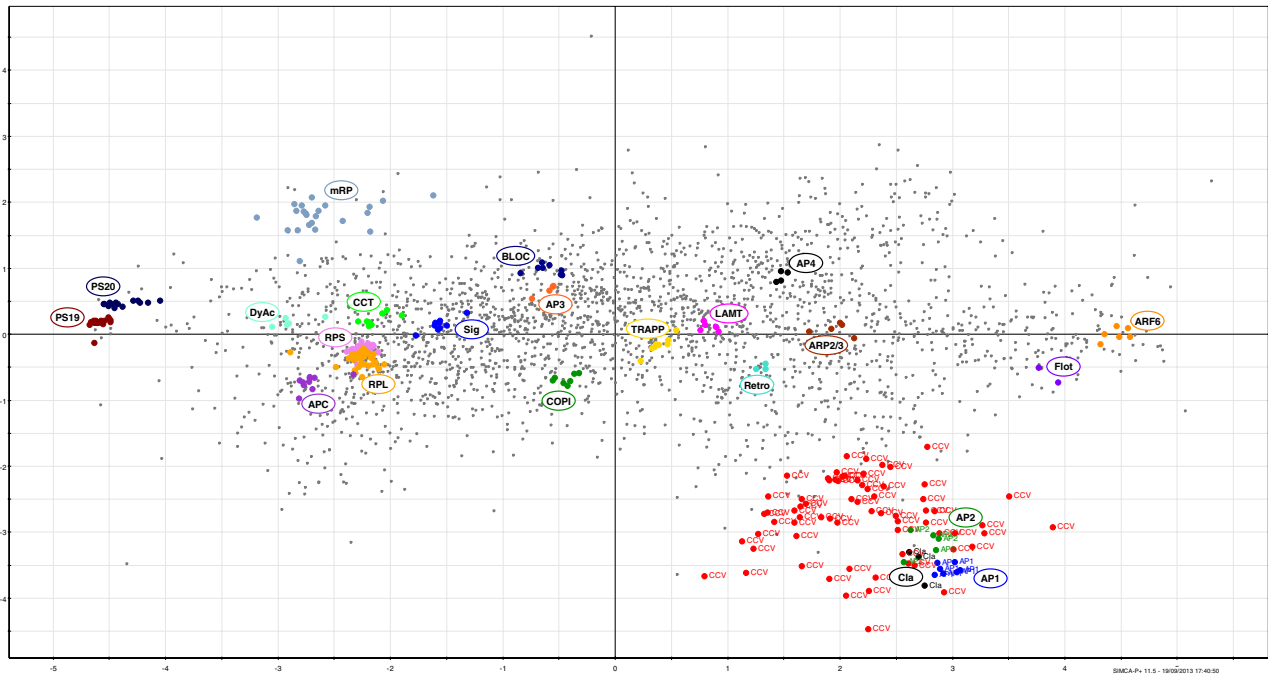


FIGURE 2: Cluster analysis of fractionation profiles accurately identifies CCV proteins and reveals functionally associated proteins in HeLa cells. The 2827 complete fractionation profiles from HeLa cells were subjected to PCA to achieve dimensionality reduction of the six-dimensional profiling data set. Projections along the first and second principal components are shown (x- and y-axis, respectively; cumulative $R^2 = 0.893$). Each protein is represented by a scatter point. Proximity of proteins indicates similar profiles and thus likely functional association. Core subunits of known complexes are highlighted in color and include AP-1 to AP-4, the anaphase-promoting complex (APC), the ARP2/3 complex (ARP2/3), the biogenesis of lysosome-related organelles complex-1 (BLOC), clathrin heavy and light chains (Cla), the T-complex (CCT), COPI (COP), the dynactin complex (DyAc), the LAMTOR complex (LAMT), the mitochondrial ribosomal large subunit (mRP), the proteasome core (PS20) and regulatory particle (PS19), Retromer (Retro), ribosomal small (RPS) and large (RPL) subunits, the signalosome (Sig), and the TRAPP complex (TRAPP). In addition, flotillins (Flot) and known cargo proteins of the ARF6-dependent endocytic pathway (ARF6) are indicated. Strikingly, in all cases, tight clustering of complex subunits is observed, suggesting that the method is capable of identifying functionally linked proteins. Known CCV proteins predicted by fractionation profiling are shown in red. They cluster in the immediate vicinity of clathrin and its adaptors (AP-1 and AP-2) and can thus be distinguished from non-CCV proteins. Hence fractionation profiling accurately predicts the composition of HeLa CCVs.

Drosophila. An extended search with lower stringency revealed up to 50 candidate CCV proteins (Supplemental Table S4). To validate some of our predictions, we tagged four candidate coat proteins (Figure 3D). All four showed extensive colocalization with established markers of CCVs. In sum, fractionation profiling was successfully implemented to characterize CCVs from S2 cells. Given the evolutionary distance between humans and *Drosophila*, it is highly likely that the approach will be applicable to a wide variety of cell types.

Fractionation profiling reveals the composition of protein complexes

Stably associated proteins are expected to show near-identical fractionation behavior and thus similar profiles; in turn, this may be exploited to predict the composition of protein complexes. To test this, we focused on the HeLa data set, as the annotation of the human protein database is much more comprehensive than that of *Drosophila*. We first investigated the well-characterized and abundant CCT chaperonin T-complex (Figure 4A). All eight

associated proteins (AP-1, AP-2, and GGA2) have equivalent profiles. Nonclathrin vesicle coats (AP-3, AP-4) and large protein particles (EF-2, ribosomal elongation factor 2) have distinct patterns. Data in B–D are in excellent agreement. (E) Fractionation profiling as a proteomic method. Fractions 1–3 and Ref1 were prepared as in A. Fractions 4–6 and Ref2 were prepared similarly but with reversed metabolic labels (heavy 4–6, light Ref2). Fractions 1–3 were pooled with Ref1 and fractions 4–6 with Ref2 and analyzed by mass spectrometry. For each identified protein, six ratios of enrichment/depletion relative to the reference fraction were obtained. (F) Proteomic profiles of clathrin (AP-1, AP-2) and nonclathrin (AP-3, AP-4) adaptor protein subunits, showing the relative abundance distribution across the six subfractions (\log_2 scale). Subunits of the same AP-complex have near-identical profiles, demonstrating the accuracy of the method. Clathrin, AP-1, and AP-2 have extremely similar profiles, suggesting that they are all part of the same subcellular structure (i.e., CCVs). In contrast, nonclathrin adaptors AP-3 and AP-4 have profiles different from clathrin and different from each other. Note that each profile consists of two independent data triplets obtained with slightly different spin parameters to improve profile discrimination (deliberately “imperfect” repeats; see *Materials and Methods*). Hence the two parts of the profiles (1–3 and 4–6) are not identical; they show similar patterns, which appear shifted on the y-axis.

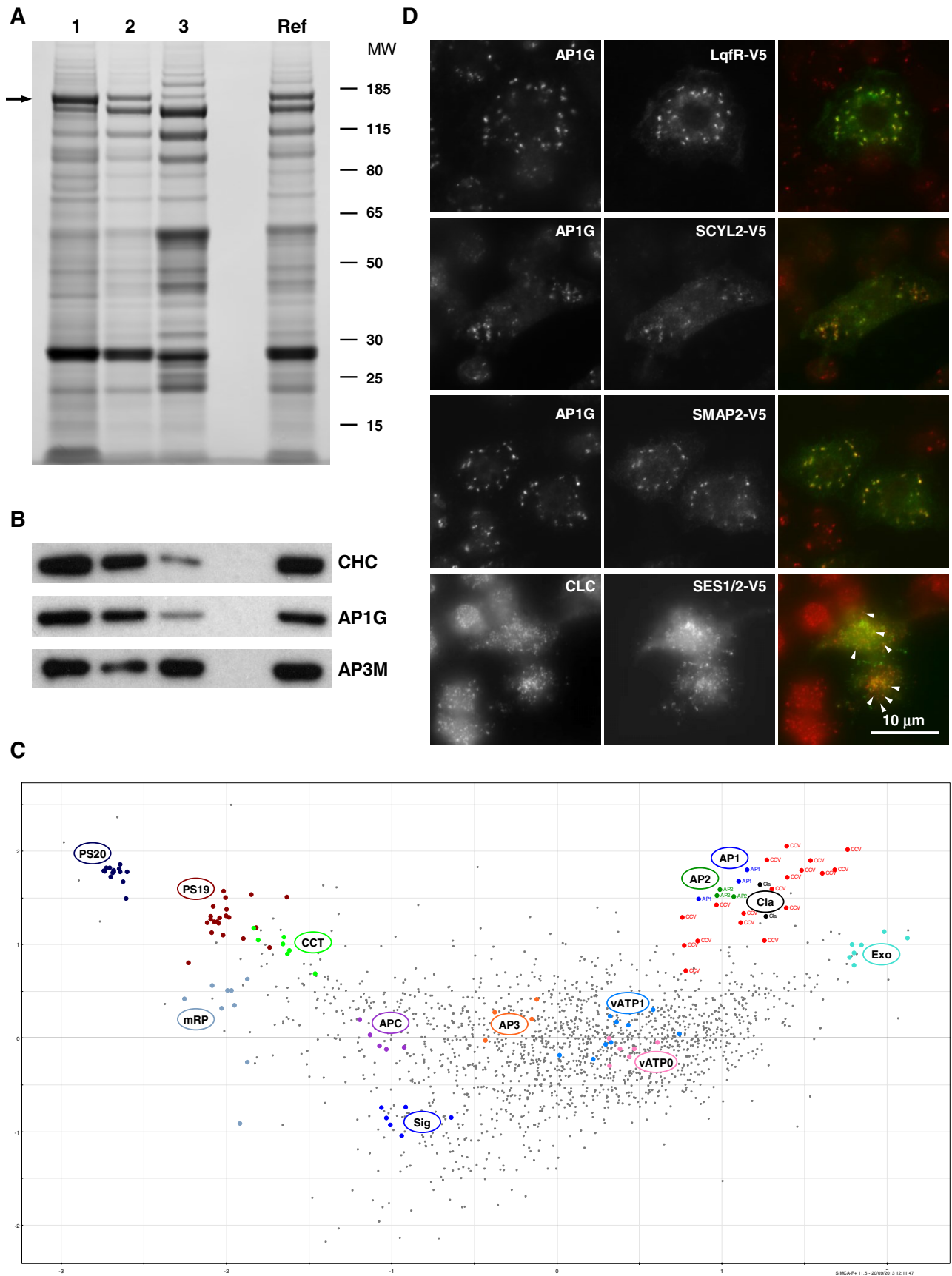


FIGURE 3: Application of fractionation profiling to S2 cells reveals the composition of *Drosophila* CCVs. (A) *Drosophila* S2 lysates were subfractionated as in Figure 1A. Gel electrophoresis (Coomassie stain) reveals that all subfractions have different compositions. The probable clathrin heavy chain band is indicated (arrow). (B) Western blotting of fractions

subunits have extremely similar profiles, and querying the Predictor with any subunit retrieves the other seven subunits as top hits. Furthermore, the mass spectrometric data allow an estimation of relative protein abundance. For the T-complex, an equimolar stoichiometry of all subunits is suggested (Figure 4B), as previously reported (Yebeles *et al.*, 2011). Hence profiling correctly predicts the composition of the T-complex *de novo*. Next we tested whether fractionation profiling can also predict the organization of more elaborate complexes. The 26S proteasome consists of two subcomplexes—the 20S core and the 19S regulatory particle. The latter is further divided into lid and base (Tomko and Hochstrasser, 2013). A search of the predictor with consensus profiles for the 19S or 20S particles identified 32 (of 33) subunits of the 26S proteasome, as well as several accessory factors. To gain further insights into complex arrangement, we applied PCA to the subunit profiles (Supplemental Figure S1A). Core and regulatory particles were clearly discriminated. Furthermore, the separation of lid and base, as well as the arrangement of core inner and outer ring, were also discernible. Finally, profiling predicts that most core and regulatory subunits are present in roughly equimolar quantities, unlike most accessory factors, which are present in substoichiometric amounts (Supplemental Figure S1B).

Encouraged by this proof of principle, we investigated the structure of less-well-studied mammalian protein complexes. TRAPP is a multisubunit vesicle-tethering complex (Yu and Liang, 2012). In yeast, the core complex TRAPPI (subunits 1–6) can associate with subunits 9 and 10 to form TRAPP II or with subunit 8 to form TRAPP III; all three TRAPPs have different functions. The mammalian TRAPP has only recently been characterized through genetic interactions and pull-down experiments (Bassik *et al.*, 2013). PCA of our profiling data clearly support the postulated existence of a mammalian TRAPP III complex consisting of subunits 1, 2, 2L, 3–5, 6A/6B, 8, and 11–13 (Figure 4C). In addition, we provide the first experimentally determined estimated TRAPP III subunit stoichiometry (Figure 4D). In agreement with the yeast model, our data suggest an equimolar stoichiometry for all core subunits 1–6, except for subunit 3, which is present in two copies. In addition, it appears that the recently identified subunits 11 and 12 are also present in single copies, with subunit 13 probably less tightly associated. Finally, the fact that subunit 8 (the defining feature of TRAPP III) is almost as abundant as the core subunits 1–6 implies that the majority of TRAPPI is part of TRAPP III and that very little TRAPPI is “free” or part of TRAPP II in HeLa cells.

The yeast Gid complex is a multisubunit E3 ligase with key regulatory functions in glucose metabolism (Menssen *et al.*, 2012). The mammalian homologue has been termed the CTLH complex (Kobayashi *et al.*, 2007), and its components have been implicated in regulation of endosomal trafficking (Heisler *et al.*, 2011) and cell spreading (Valiyaveetil *et al.*, 2008). Pull-down experiments have identified six members of the CTLH complex: ARMC8, C20orf11, MAEA, MKLN1, RANBP9, and RMND5 (Kobayashi *et al.*, 2007). Our profiling data support this composition (Figure 4, E and F) but clearly indicate the presence of three further subunits: C17orf39 and WDR26, the human homologues of yeast GID4 and GID7, respectively; and VPRBP (the HIV-VPR binding protein), which has no homologue in yeast. VPRBP is a particularly intriguing new discovery. This protein associates as a substrate adaptor with the CRL4A E3 ligase CUL4A/DBB1/RBX1 (Romani and Cohen, 2012). The CRL4A complex is also present in our profiling data, but its profile is distinct from that of the CTLH complex (Figure 4, G and H). Furthermore, VPRBP has a LisH protein interaction domain, which is also present in several CTLH subunits (Menssen *et al.*, 2012). Our data therefore suggest that VPRBP may have an additional function as a substrate adaptor in CTLH. Of importance, the HIV protein VPR is known to “highjack” VPRBP and its associated E3 ligase to trigger degradation of as-yet-unidentified targets (Romani and Cohen, 2012). We propose that the CTLH complex may also become a target of HIV-VPR and that it may play a role in HIV pathogenesis.

Fractionation profiling predicts the existence of novel protein complexes and trafficking pathways

The foregoing data demonstrate the predictive power of our approach for known protein complexes. However, profiling also predicts the existence of novel complexes (Figure 5, Supplemental Figure S2, and Supplemental Table S5). For example, we provide the first evidence that the BAR-domain proteins SNX4 and SNX30 (van Weering *et al.*, 2010) form a stable dimer (Figure 5, A and B). A recent study identified a novel protein dimer C17orf75/WDR11, implicated in protection against ricin toxicity (Bassik *et al.*, 2013). Our profiling data suggest that these proteins are in fact part of a trimeric complex, which also includes the protein FAM91A1 (Figure 5, C and D). The protein C10orf32 has been implicated in susceptibility to arsenic poisoning (Pierce *et al.*, 2012); here we propose that C10orf32 is in complex with the uncharacterized protein LOIH12CR1 (Figure 5, E and F). These examples illustrate that our approach covers a broad spectrum of important novel protein associations.

shown in A confirms that CHC and the associated adaptor AP-1 have similar profiles, which are distinct from the nonclathrin adaptor AP-3. (C) Proteomic analysis of S2 fractions from two independent profiling experiments identified 1799 proteins with complete profiles. PCA of the profiles reveals clustering of subunits of known protein complexes, including AP-1, AP-2, AP-3, the anaphase-promoting complex (APC), the T-complex (CCT), clathrin heavy and light chains (Cla), the Exocyst (Exo), the mitochondrial ribosome (mRP), the proteasome core (PS20) and regulatory particle (PS19), signalosome (Sig), and the V-ATPase (vATP0, integral membrane subcomplex; vATP1, peripheral subcomplex). Of importance, proteins whose mammalian homologues are known CCV proteins cluster in the vicinity of clathrin (marked in red, CCV). Fractionation profiling successfully reveals the composition of the *Drosophila* CCV proteome. (*x*, *y*-axes = first and second principal components; cumulative $R^2 = 0.917$). To illustrate that even a single fractionation profiling experiment is sufficient to produce meaningful clustering, this plot shows PCA of the first triplet only (i.e., three data points). Joined PCA of both triplets (all six data points) results in a very similar plot. (D) Four candidate *Drosophila* CCV coat components predicted by fractionation profiling were C-terminally tagged and transiently expressed in S2 cells. Proteins were visualized by immunofluorescence microscopy. All four proteins colocalize with established markers of clathrin-coated vesicles, confirming their association with CCVs in S2 cells. LqfR, SCYL2, and SMAP2 colocalize tightly with AP-1, a marker of TGN/endosomal CCVs. SES1/2 localizes to plasma membrane clathrin-coated pits. Because clathrin light chain (CLC) marks both endocytic and intracellular clathrin-coated structure, it shows only partial colocalization with SES1/2. Some areas of colocalization are indicated (white arrowheads). Scale bar, 10 μ m.

Rank	<i>Drosophila</i> gene (UniProt)	Protein ID (UniProt)	Relative abundance	Predicted role	Human homologue	Known hCCV protein?
1	Chc	P29742	1,000,000	Coat	CLTC	Yes
2	Clc	E1JI22	1,042,971	Coat	CLTA/B	Yes
3	Ack	Q9VZ12	13,163	Coat	TNK2	Yes
4	lqfR ^a	Q8IN05	9986	Coat	CLINT1/EpsinR	Yes
5	Rabex-5	Q9W0H9	1435	Coat	RABGEF1	Yes
6	Ocrl	O46094	11,232	Coat	OCRL	Yes
7	AP- γ	Q9W388	128,156	Coat	AP1G1	Yes
8	Bap	Q24253	125,375	Coat	AP1B1/AP2B1	Yes
9	PI3K68D	Q7K3H0	21,936	Coat	PIK3C2A	Yes
10	Fur1	P30430	916	Cargo	FURIN	Yes
11	AP-50	O62530	59,444	Coat	AP2M1	Yes
12	AP-47	O62531	151,582	Coat	AP1M1	Yes
13	Aux	Q9VMY8	38,059	Coat	DNAJC6/auxilin	Yes
14	Lqf	Q9VS85	8993	Coat	Epsin1/2/3	Yes
15	CG2747	Q7KSW1	4262	Coat	HEATR5B/p200	Yes
16	AP-2 σ	Q9VDC3	39,042	Coat	AP2S1	Yes
17	Lerp	B7Z0Q8	12,518	Cargo	CIMPR	Yes
18	α -Adaptin	P91926	69,120	Coat	AP2A1/2	Yes
19	Syx6	Q7JY00	2707	Cargo	STX6	Yes
20	CG7800-RA	Q9VHV3	214,353	Cargo	—	—
21	Gga	Q9W329	13,059	Coat	GGA1/2/3	Yes
22	CG8243-RA ^a	A1Z7K6	4122	Coat	SMAP2	Yes
23	CG12393 ^a	B7Z022	7991	Coat	SES1/2	Yes
24	Syx16	Q9VR90	11,908	Cargo	STX16	Yes
25	CG1951 ^a	Q9VAR0	14,984	Coat	SCYL2/CVAK104	Yes
26	AP-1 σ	Q9VCF4	20,672	Coat	AP1S1/2/3	Yes
27	svr	P42787	26,406	Cargo	CPD	Yes
28	Aut1	Q9VVS6	2200	Coat	ATG3	—
29	Vps45	Q9VHB5	26,257	Coat	VPS45	Yes

The *Drosophila* S2 cell Predictor (Supplemental Table S3) was searched against clathrin heavy chain (Chc). Predictions are ranked in order of similarity to Chc. Only the predictions with the highest level of confidence and a minimum of five data points are shown here (see Supplemental Table S4 for the complete list of predicted *Drosophila* S2 CCV proteins). Relative abundance is indicated in arbitrary units. The last column indicates whether the human homologue of an identified protein is a known CCV constituent.

^aProtein subcellular localization investigated by immunofluorescence microscopy (Figure 3D).

TABLE 1: The *Drosophila* S2 CCV proteome: top predictions.

Profiling also makes predictions about the subcellular distribution patterns of integral membrane proteins. In this case, similar profiles suggest similar trafficking itineraries. As an example, we investigated the ARF6-dependent endocytic pathway. Eyster *et al.* (2009) identified a group of four plasma membrane proteins (CD44, CD98 [SLC3A2/SLC7A5], CD147 [BSG], and ICAM1) that are endocytosed in a clathrin-independent manner and follow an ARF6-dependent recycling route. Indeed, these proteins have extremely similar profiles (Figure 5G), to the extent that a search with CD44 identifies all other three proteins as top hits. In addition, this cluster includes the monocarboxylate transporter (SLC16A1/MCT1), a key regulator of nutrient uptake, with a proposed role in tumor proliferation (Pinheiro *et al.*, 2012). We predict that this protein follows the same endocytic and ARF6-dependent recycling

route. Finally, we investigated whether this pathway is related to flotillin- or caveolin-mediated endocytosis (Sandvig *et al.*, 2011). PCA shows that profiles of these markers of clathrin-independent endocytic pathways are clearly different from those for the ARF6 cargo cluster (Figure 5H).

DISCUSSION

Here we developed proteomic fractionation profiling, a simple yet powerful tool for mapping functional protein associations. We have mainly used fractionation profiling for the analysis of CCVs and protein complexes, but the method is likely to work equally well for the characterization of other types of vesicular transport intermediates. Our results are presented in the interactive Predictor database (Tables S1 and S3).

Fractionation profiling for the analysis of vesicle composition

Fractionation profiling is based on identifying proteins with similar subcellular distributions. This concept has been exploited in several published methods, including localization of organelle proteins by isotope tagging (LOPIT; Dunkley *et al.*, 2004) and protein correlation profiling (PCP; Foster *et al.*, 2006). Both approaches rely on density gradient centrifugation to separate organelles; quantification of protein distribution across the gradient is achieved by isotope tagging (LOPIT) or label-free quantification (PCP). Although these methods are conceptually powerful, they necessitate large numbers of fractions, long measuring times, and complex data analysis. Crucially, neither method was able to resolve transport vesicles, such as CCVs. Fractionation profiling pairs the superior accuracy of SILAC quantification with the robustness and simplicity of differential centrifugation. It requires minimal amounts of starting material and instrument time and is computationally straightforward. We show that fractionation profiling can determine the composition of CCVs from HeLa cells with excellent sensitivity and specificity, comparable to our perturbation-based multivariate profiling approach (Borner *et al.*, 2012). Unlike our previous method, however, fractionation profiling does not involve the use of small interfering RNA knockdown, which is a distinct advantage, since not all cell types are amenable to knockdown. To demonstrate that the approach is not limited to HeLa cells, we applied it to evolutionarily distant *Drosophila* S2 cells. Without requiring any cell-specific optimization, fractionation profiling provided the first comprehensive insect CCV proteome. Furthermore, our data also allow the first specific interspecies comparison of CCV content and thus identify evolutionarily conserved core machinery and cargo (Table 1).

Our findings suggest that the approach will be applicable to most cell types amenable to SILAC labeling and thus provide a long-awaited universal tool for characterizing the composition of CCVs. Additional strengths of the method are its relative simplicity and rapid yield of results. In our experience, even a single fraction triplet already generates a very accurate draft CCV proteome (illustrated in Figure 3C). An exciting prospect is the application of the method to unicellular eukaryotes. For example, trypanosome pathogenicity depends on clathrin-mediated endocytosis (Allen *et al.*, 2003), yet the trypanosome CCV composition remains largely unknown. Fractionation profiling could quickly provide answers and thus identify new drug targets for treatment of sleeping sickness. We expect that fractionation profiling has the potential to become a standard tool for the proteomic analysis of CCVs.

Fractionation profiling as a tool for mapping protein complexes

A second major result of this work is that fractionation profiling allows the characterization of protein complexes. Several previous studies used comparative approaches to investigate protein interaction networks. For example, Havugimana *et al.* (2012) analyzed cell lysates by multiple chromatographic techniques. The distribution of 3000 proteins was quantified by mass spectrometry across >1000 subfractions, and 622 putative complexes were identified (from ~14,000 binary interactions). Similarly, Kristensen *et al.* (2012) used size-exclusion chromatography to separate cell lysates into 50–100 fractions and applied protein correlation profiling (Foster *et al.*, 2006) to identify 291 complexes from 3400 quantified proteins (and ~7000 binary interactions). Although both studies yielded impressive catalogues, they are largely restricted to cytosolic protein complexes, as the used separation techniques cannot easily cope with membrane fractions. Furthermore, both methods require substan-

tial mass spectrometric analysis time, owing to the large number of fractions. Here we introduced a comparatively simple method—fractionation profiling—which requires analysis of only six fractions, yet provides very accurate predictions. Furthermore, the method is particularly suitable for the analysis of membrane-associated protein complexes. The HeLa data cover >3500 protein profiles, approximately one-third of the HeLa proteome (Nagaraj *et al.*, 2011). The Predictor does not rigidly assign complex boundaries. However, we identified 6980 pairs of profiles with “very high similarity,” which are comparable to predicted binary interactions. We therefore estimate that at least several hundred protein complexes are covered by our data set. These include large cytosolic particles such as ribosomes and proteasomes (megadalton range), but also smaller assemblies, such as the CTLH complex (~600 kDa). Very small cytosolic complexes are probably not pelleted under the conditions used here; however, complexes associated with membranes are very well resolved, regardless of size. Even small dimers (such as SNX4/SNX30; ~100 kDa) are readily identified. Indeed, our analysis includes numerous membrane-interacting complexes that not represented in either Havugimana’s (2012) or Kristensen’s (2012) study, such as AP-4, BLOC-1, CTLH, LAMTOR, and TRAPP. In addition, our method predicts groups of integral membrane proteins with similar trafficking pathways, such as ARF6-dependent endocytic cargo proteins, and thus goes beyond the analysis of complexes.

The information gleaned from a profiling analysis is threefold. First, searching the Predictor against a protein of interest identifies candidate functionally associated proteins. The Predictor provides a classified output, which allows users to evaluate the relevance of the retrieved proteins. The Predictor can hence identify novel complexes or candidate novel subunits of known complexes. Second, the Predictor provides estimated relative abundances of the retrieved proteins. We demonstrated the validity of this feature by correctly predicting the stoichiometry of numerous known complexes (Figure 4 and Supplemental Figure S1). Third, the retrieved profiles can be further analyzed by PCA to give insights into the structural organization of protein complexes. To test this, we first predicted the subcomplex arrangement of the 26S proteasome (Supplemental Figure S1A); we then used the method to define the composition of the HeLa TRAPPIII complex (Figure 4C) and also to map the association of VPRBP with CTLH versus the CRL4A E3 ligase (Figure 4H). To our knowledge, there is no published method as simple as fractionation profiling that grants similarly detailed insights into protein complex composition and organization.

Conclusion

Here we exploited fractionation profiling as a tool for characterizing CCV composition. We also used it to analyze a number of known and uncharacterized protein complexes to illustrate the power of the approach. These examples represent only a small proportion of the information in the profiling data sets. We therefore present the complete data as a community resource (the Predictor, Supplemental Tables S1 and S3). Readers are strongly encouraged to search the database against proteins of interest. The Predictor is very easy to use; a quick-start guide is included in the file, and a detailed manual is provided separately (Supplemental Predictor Manual).

Finally, we propose that fractionation profiling can be adapted to other cell biological questions. In principle, any homogeneous subcellular structure should be amenable to a profiling analysis, provided it is possible to prepare an enriched fraction as the starting point. Good candidate targets include, for example, neuronal secretory granules (Bonnemaison *et al.*, 2013) and the elusive “Glut4 storage vesicles” in adipocytes (Leto and Saltiel, 2012).

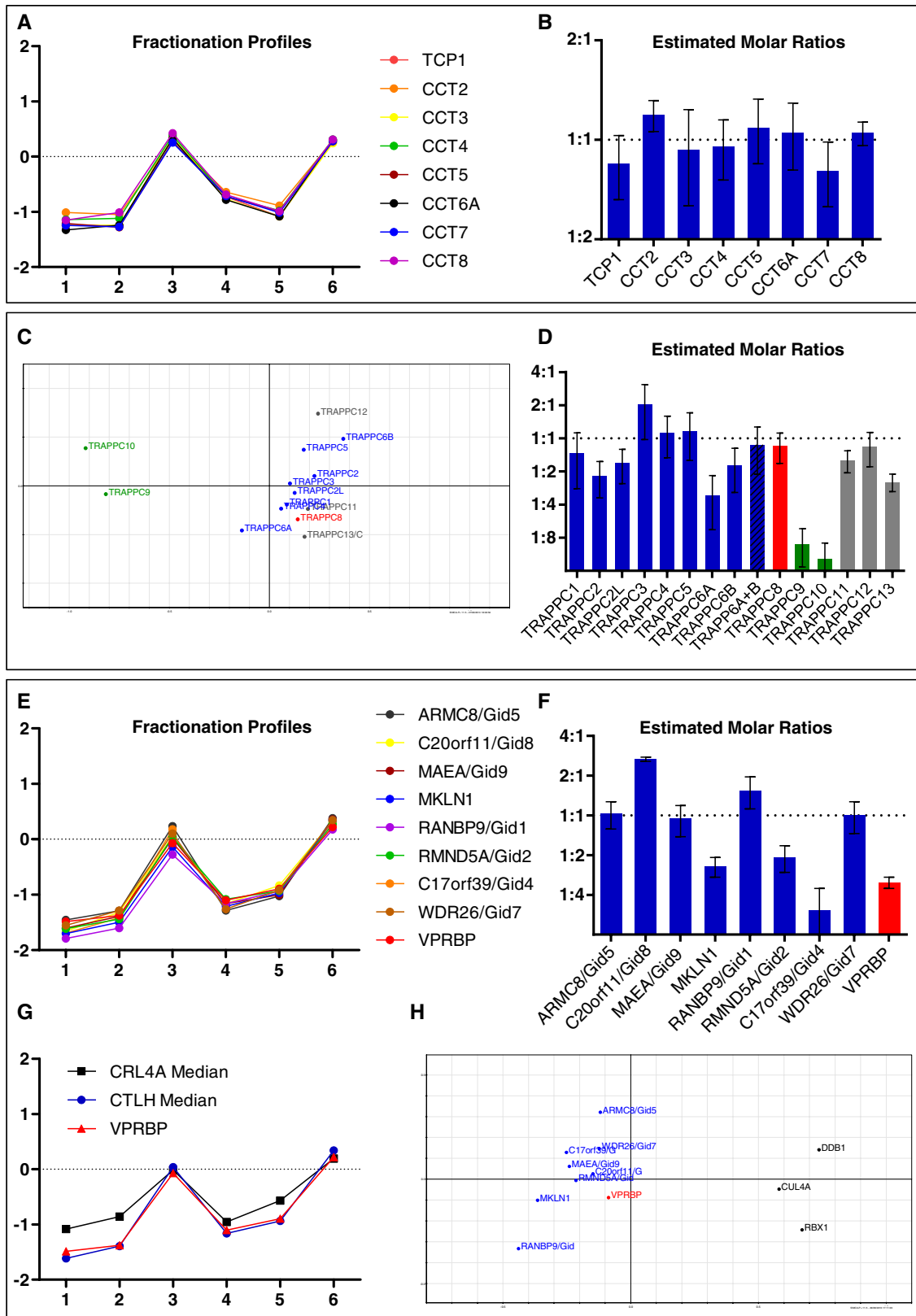


FIGURE 4: Fractionation profiling predicts the composition and stoichiometry of protein complexes. (A, B) The chaperonin T-complex (eight subunits). (A) Fractionation profiles, showing the protein abundance in each fraction relative to the reference (log₂ scale). Profiles of individual T-complex subunits are nearly superimposable. (B) Estimated stoichiometry, normalized to the median of all subunits. Fractionation profiling correctly predicts the equimolar stoichiometry of the complex. (C, D) The TRAPP complex (14 subunits). Yeast TRAPP exists in three configurations: the TRAPPI core (1, 2, 2L, 3–5, 6A, and 6B) can associate either with subunits 9 and 10 to form TRAPPII or with subunit 8 to form TRAPPIII. (C) PCA of subunit profiles reveals that in HeLa cells, the majority of TRAPP is assembled into

Conversely, the approach may be focused primarily on protein complexes, for example, by subfractionating nuclear or mitochondrial extracts. The transfer to other biological targets will probably require some optimization of the fractionation conditions. Nevertheless, the experimental design and subsequent data analysis can be modeled on the approach presented here, as discussed further in a dedicated section at the end of *Materials and Methods*. Fractionation profiling is an extremely versatile yet conceptually straightforward tool for mapping functional protein interactions, with many potential future applications beyond those demonstrated here.

MATERIALS AND METHODS

Fractionation profiling

The following protocol describes the procedure as used in this study for the analysis of CCVs. The final section discusses conceptual aspects of the approach, limitations, and caveats, as well as considerations for future applications.

Overview

A fractionation profiling experiment consists of two parallel sample preparations: a "reference" sample (e.g., from SILAC heavy-labeled cells) and a set of "subfractions" (e.g., prepared from SILAC light-labeled cells). Labels are usually swapped for a repeat experiment. The first part of the protocol follows the method for preparing a CCV-enriched fraction described in Borner *et al.* (2012).

Preparation of cell lysates

All operations were performed at 4°C. Adherent HeLaM cells (recommended 1000 cm² confluent/lysate) were washed once in ice-cold phosphate-buffered saline (PBS) and once in ice-cold MES buffer (0.1 M 2-(*N*-morpholino)ethanesulfonic acid [MES], pH 6.5 [adjusted with NaOH], 0.2 mM ethylene glycol tetraacetic acid, and 0.5 mM MgCl₂). Cells were drained carefully and scraped into 10 ml of ice-cold MES buffer. Cells were lysed in a 30-ml Potter-Elvehjem homogenizer (clearance ~100–150 μm) with 20 strokes of a motorized pestle.

Suspension *Drosophila* S2 cells (recommended ~0.5–1 ml of pelleted cells/lysate) were washed once in ice-cold PBS and once in ice-cold MES buffer. Cells were resuspended in 6 ml of ice-cold MES buffer and first lysed in a Potter-Elvehjem homogenizer as described. To improve the lysis of the relatively small S2 cells, primary lysates were additionally passed through a 21-gauge (0.8-mm) needle eight times.

Subcellular fractionation

All operations were performed at 4°C. Lysates were centrifuged at 4000 × *g* for 32 min to pellet unbroken cells and large debris. Supernatants were treated with 50 μg/ml ribonuclease A (≥2.5 Kunitz units/ml) for 1 h, and partially digested ribosomes were pelleted by

centrifugation at 4000 × *g* for 3 min. Supernatants were then centrifuged at 209,000 × *g* (55,000 rpm, MLA-80 rotor; Beckman Coulter, High Wycombe, UK) for 40 min to pellet membranes. Pellets were resuspended in 400 μl of MES buffer with 60 strokes in a 1-ml Wheaton homogenizer (tight pestle, clearance ~25–75 μm). An equal volume of F/S buffer (12.5% [wt/vol] Ficoll [PM400; GE Life Sciences, Little Chalfont, UK] and 12.5% [wt/vol] sucrose in MES buffer) was added and carefully mixed, and samples were centrifuged at 21,700 × *g* (20,000 rpm, TLA-110; Beckman Coulter) for 34 min. Supernatants were diluted 1:5 in MES buffer. Up to this stage, the preparation procedure is identical for the (SILAC heavy) reference and (SILAC light) subfraction samples (and essentially as described in Borner *et al.*, 2012).

1. For the preparation of the "reference" fraction, the diluted supernatant was centrifuged at 195,500 × *g* (60,000 rpm, TLA-110) for 40 min to yield the "reference pellet."
2. For the generation of "subfractions," the diluted supernatant was first centrifuged at 66,500 × *g* (35,000 rpm, TLA-110) for 20 min to yield the 35K pellet. The supernatant was transferred to a fresh centrifuge tube, mixed by pipetting, and centrifuged at 110,000 × *g* (45,000 rpm, TLA-110) for 20 min to yield the 45K pellet. The supernatant was again transferred to a fresh tube and centrifuged at 195,500 × *g* (60,000 rpm, TLA-110) for 40 min to obtain the final 60K pellet.

The spin parameters of the reference pellet correspond to those of the 60K subfraction pellet. Hence the contents of the reference pellet correspond to those of all three subfraction pellets combined (35K + 45K + 60K).

The recommended procedure for sample resuspension is to resuspend all pellets in 1× sample buffer (made from NuPAGE 4× LDS sample buffer [Life Technologies, Paisley, UK] diluted with water). Each subfraction pellet should be redissolved in 20 μl and the reference pellet in 60 μl. This method is particularly suitable for small sample yields (e.g., if limited starting material is available). Alternatively, samples may be dissolved in a suitable volume of SDS buffer (2.5% [wt/vol] SDS, 50 mM Tris-HCl, pH 8) for estimation of protein yields before gel electrophoresis.

Note that for HeLa cells, the two repeats of the fractionation profiling were deliberately performed with slightly different spin parameters (35K, 45K, and 60K, first triplet; 30K, 40K, and 60K, second triplet). The profiles obtained with the different sets of spins are correlated but appear somewhat shifted on the *y*-axis in Figure 1F. The intention behind this "imperfect" repeat was to enhance the resolution of subcellular structures by teasing out more differences in fractionation behavior. Inspection of the PCA loadings plot corresponding to Figure 2, however, indicates that both sets of spins make very similar contributions to the analysis. The gain in resolution is thus likely to be relatively small. For the S2 cells, we used identical spin conditions for both repeats. In our experience, either method performs well.

TRAPPIII. Subunits 11–13 are also part of this complex. Core subunits are shown in blue, the TRAPPIII subunit 8 is shown in red, the TRAPPII subunits 9 and 10 are shown in green, and subunits 11–13 are shown in gray. (D) Estimated stoichiometry of mammalian TRAPP. (E–H) The CTLH complex (six known plus three novel subunits). Yeast homologues of mammalian proteins are indicated (GID; Menssen *et al.*, 2012). (E) Subunit profiles. Three candidate novel subunits—C17orf39, WDR26, and VPRBP—were identified. (F) Estimated stoichiometry of the mammalian CTLH complex predicts a core of ARMC8, C20orf11, MAEA, RanBP9, and WDR26, with Muskelein, C17orf39, RMND5a, and VPRBP possibly less stably associated. (G) VPRBP has been proposed to function as a substrate adaptor in the E3 ligase CRL4A (Romani and Cohen, 2012). The CRL4A ligase consist of three proteins (CUL4A, DDB1, RBX1), which are also present in the profiling set. Consensus (median) profiles of CRL4A and CTLH were plotted. The profile of VPRBP matches that of CTLH. (H) PCA of subunit profiles confirms that VPRBP (red) profiles closely with CTLH subunits (blue) and not with CRL4A subunits (black). Error bars of abundance scores correspond to MADs; *n* = 6 replicate measurements.

Sample preparation for mass spectrometric analysis

Two alternative protocols were used for sample mixing:

1. Equal-proportions mixing (recommended): Each subfraction sample (35K, 45K, and 60K, all SILAC light labeled if the reference is SILAC heavy) was combined with one-third of the reference sample. Using the recommended resuspension volumes, each of the three samples should be ~40 μ l. Two microliters of 1 M dithiothreitol (DTT) was added to each sample (50 mM final), and samples were incubated at 90°C for 3 min.
2. Equal-protein mixing: An alternative protocol was to estimate the protein concentration in each fraction by bicinchoninic acid assay (Thermo Fisher Scientific, Cramlington, UK) and mix equal protein quantities of reference and subfractions. This method optimizes protein load, but a proportion of the sample is lost for the protein assay. It is hence recommended only for preparations with large yields. DTT was added from a 1 M stock to a final concentration of 50 mM, and samples were incubated at 90°C for 3 min.

The two methods result in slightly different ratios; it is hence recommended to use the same method between repeat fractionation profiling experiments to ensure reproducibility of profiles.

Mass spectrometry analysis and raw data processing

Samples were separated by one-dimensional gel electrophoresis, using precast gradient gels (NuPAGE; Invitrogen). Gels were stained with colloidal Coomassie, and each lane was cut into 10 pieces. A typical fractionation profiling experiment thus comprised 30 gel slices, which were subjected to tryptic digest (Shevchenko *et al.*, 2006). Peptide extracts were cleaned up and stored at 4°C on StageTips (Rappsilber *et al.*, 2007). Samples were eluted from StageTips, separated on a 20-cm reverse-phase column, and analyzed on a Q Exactive mass spectrometer (Thermo Fisher Scientific), essentially as described (Nagaraj *et al.*, 2012).

Each fractionation profiling experiment included two triplets of samples; each sample consisted of 10 gel slices/mass spectrometry runs. All (60) raw files were jointly processed with MaxQuant, version 1.3.0.5 (Cox and Mann, 2008), and its built-in Andromeda search engine (Cox *et al.*, 2011), using UniProt (www.uniprot.org) reference databases for human or *Drosophila* proteins. Cam (C) was set as a fixed modification and oxidized (M) and acetyl (Protein N-terminus) as variable modifications. Both peptide and protein false discovery rates were set to 0.01; the minimum peptide length was seven amino acids. Calculations of ratios were performed on razor and unique peptides with the requantify and match-between runs features enabled. The minimum ratio count was two.

Data analysis

The protein identification lists from MaxQuant (Cox and Mann, 2008) were filtered by removing matches to the reverse database, proteins only identified with modified peptides, and common con-

taminants. Abundance ratios were then \log_2 transformed. For HeLa cells, we identified 4523 proteins (minimum, one ratio in the six samples); 3549 had at least one complete fractionation triplet (i.e., a usable profile), and 2827 had complete profiles (six ratios). For S2 cells, the corresponding numbers were 3163 (minimum, one ratio), 2483 (minimum, one data triplet), and 1799 (complete profile). All identifications are presented in Supplemental Tables S1 (HeLa) and S3 (*Drosophila* S2).

PCA was performed as described in Borner *et al.* (2012), using SIMCA-P+ 11.5 (Umetrics, Crewe, UK). Scatter plots in Figures 2–5 and Supplemental Figures S1 and S2 were also generated in SIMCA-P+ and further annotated in PowerPoint (Microsoft, Reading, UK). Line and column plots shown in Figures 1, 4, and 5 and Supplemental Figure S1 were prepared within Prism 6 (GraphPad Software, La Jolla, CA) and further annotated in PowerPoint (Microsoft).

Construction of the Predictor database

The Predictor provides a very simple interface that allows users to query the profiling data (Supplemental Tables S1 and S3). Its purpose is to predict groups of functionally associated proteins on the basis of common subcellular fractionation behavior. Using the Predictor requires only the most basic familiarity with Excel (Microsoft). The Predictor contains all proteomic data presented in this study in a compact yet highly accessible format. A quick-start guide is included in the same file. In addition, a detailed manual, including guidance to the interpretation of results, is provided as a separate document (Supplemental Predictor Manual).

Overview

Abundance profiles were tabulated in the raw data tables (Supplemental Tables S1 and S3; see Complete Data tab). Using standard Excel spreadsheet functions, an interactive interface (the Predictor) was created, which allows users to input one or several query genes and specify various search parameters. The Predictor then retrieves the corresponding profiles from the raw data table and calculates a consensus query profile. This query profile is automatically compared with all other profiles in the raw data table. Profiles are then sorted by similarity to the query profile, as shown in the results table. In addition, the Predictor provides various graphical summaries for both query and output profiles, including estimates of relative protein abundance. The details of the calculations are provided in what follows.

Distance calculations and ranking

A complete profile consists of six (\log -transformed) abundance ratios, R1–R6, which indicate the protein's distribution across different subcellular fractions. Proteins with similar profiles have similar subcellular distributions. To evaluate profile similarities, users can choose between two alternative measures of profile distance. For a given query profile, the distance to any other Protein X in the database is calculated as follows:

and WDR11; profiling clearly indicates a stable complex that includes FAM91A. (C) Profiles. (D) Estimated stoichiometry. (E, F) The LOH12CR1/C10orf32 dimer. (E) Profiles. (F) Estimated stoichiometry. (G, H) Cargo of the ARF6-dependent endocytic pathway. CD44, CD98 (SLC3A2 heavy chain, SLC7A5 light chain), ICAM1, and BSG (CD147) are known cargo proteins of a clathrin-independent endocytic pathway. (G) Profiles of these proteins are almost identical, confirming their similar trafficking itineraries. A further protein, SLC16A1 (monocarboxylate transporter 1), is identified as part of the same cluster. (H) PCA of individual profiles shows that the ARF6-cargo cluster (red) is distinct from flotillins (blue) and caveolins (green). SLC16A1 (black) is part of the ARF6 cluster. Error bars of abundance scores correspond to MADs ($n = 6$ replicate measurements).

1. Average absolute profile distance (“Manhattan distance”):

$$\text{Av abs distance} = [\text{ABS}(R1_{\text{Query}} - R1_{\text{Protein } \chi}) + \text{ABS}(R2_{\text{Query}} - R2_{\text{Protein } \chi}) + \dots + \text{ABS}(R6_{\text{Query}} - R6_{\text{Protein } \chi})]/6$$
2. Average squared profile distance (the average “squared Euclidean distance”):

$$\text{Av squared dis} = [(R1_{\text{Query}} - R1_{\text{Protein } \chi})^2 + (R2_{\text{Query}} - R2_{\text{Protein } \chi})^2 + \dots + (R6_{\text{Query}} - R6_{\text{Protein } \chi})^2]/6$$

Both methods produce distance values ≥ 0 . However, they produce slightly different hierarchies within the output. The squared-distance method is more stringent; because all differences are squared, large deviations from the query profile are accentuated much more than with the absolute-distance calculation. Conversely, the squared distance is more likely to produce false negatives, since a single inaccurate ratio can “distort” the whole profile similarity.

The predictor also allows to search the database with and for “incomplete” profiles, that is, profiles with fewer than six abundance ratios. In such cases, only the differences of available ratios are summed and divided by the number of available data points. Distances are thus normalized to the number of available data points, and these “average” distances are comparable between complete and incomplete profiles. To maintain high predictive power, however, it is recommended that query and output profiles should contain at least one complete fraction triplet (see earlier discussion).

Once the predictor has calculated the distance of the query profile to every protein in the database, proteins are ranked according to similarity with the query (hierarchical output).

Note that the Predictor’s distance calculations do not involve prior PCA. The PCA plots in Figures 2–5 and Supplemental Figures S1 and S2 represent graphical illustrations of profile clustering but are not required for the Predictor.

Similarity evaluation

In addition to the hierarchical output, the absolute distances to the query are also provided, together with a simple “star” system to guide the interpretation of the results (e.g., three stars denote proteins whose profile is almost identical to the query; two stars denote a useful cut-off for proteins commonly found in a complex, etc.). To derive these qualifiers, we scored distances of subunits within well-established mammalian protein complexes. Guided by the CORUM database (Ruepp *et al.*, 2010), we assembled a calibration set of known protein complexes represented in our HeLa fractionation profiling data set. We included only complexes with at least four subunits with complete profiles. Our reference set included 21 complexes, with 232 proteins in total (Supplemental Table S6). PCA showed that the subunits of most calibration complexes cluster tightly. Therefore subunits within a calibration cluster have mostly very similar profiles. For each of the 21 complexes, we calculated a consensus “center” (i.e., median) profile and determined the distance of each subunit to the center of its complex. The resulting 232 distances are representative of typical profile distances encountered in known stable protein complexes. Distances were sorted from highest to lowest, and distance values were identified to include the top 40% (almost identical; three stars), 80% (very similar; two stars), 90% (similar; one star), or 95% (borderline similar, annotated as “B”) of all calibration profiles. Cut-offs were determined for both distance scoring methods (see earlier description). Similarity filters are automatically applied to the Predictor’s output and included in the annotation.

Because the CORUM database (Ruepp *et al.*, 2010) covers only mammalian proteins, a proxy was devised for the *Drosophila* data set, using the information derived from the HeLa data set. First, a distance matrix of all complete HeLa profiles was generated

(2827 \times 2827 profiles). The corresponding 3,994,551 distances were sorted from highest to lowest. These calculations were performed in the Perseus module of the MaxQuant software suite (www.perseus-framework.org) using a custom plug-in, which is available on request. In this ranked list of distances, we located the distances corresponding to the star rankings determined as described and converted them into proportions (e.g., only the top 989 distances were smaller than the three-star cut-off, corresponding to the top 0.0247%). The analogous distance matrix was then calculated for the S2 data set (1799 \times 1799 profiles = 1,617,301 distances), and distances were again sorted from lowest to highest. The percentage cut-offs determined for the HeLa set were then applied to the S2 set to locate the corresponding distance cut-offs (e.g., 0.0247% [the three-star percentage cut-off in the HeLa set] of 1,617,301 corresponds to the top 400 distances; hence the distance of entry 400 specifies the three-star distance cut-off for the S2 set).

CCV scoring

A particular strength of the Predictor is to identify proteins associated with CCVs. The indispensable core component of all CCVs is clathrin heavy chain (commonly known as CHC, but CLTC in UniProt nomenclature). Hence all proteins with profiles similar to CLTC are candidate CCV proteins. Although all CCV proteins occupy the same subcellular compartment, the profile distances between known CCV proteins are still somewhat larger than those encountered in tight CORUM protein complexes, and a different scoring method was devised to account for this. The HeLa CCV proteome is well characterized, with 151 predicted constituents (Borner *et al.*, 2012; Hirst *et al.*, 2012). Of these, 140 were represented with complete profiles in our analysis. For each CCV protein, we calculated the profile distance to CHC and sorted distances from lowest to highest. We then determined the distance cut-offs that include the top 40% of known CCV proteins (three-star similarity), top 60% (two stars), top 80% (one star), and top 90% (B, borderline similarity). Classifiers are automatically applied to the predictor’s output and included in the annotation. For example, the predictor classifies 93 proteins as two- or three-star CCV proteins, of which 84 (60% of 140) are in the reference data set of 140.

Because no suitable reference database of fly CCV proteins was available, a simple proxy was used to determine similarity guides. We assumed that the S2 data set contains roughly the same proportion of CCV proteins as the HeLa set, scaled down by the lower complexity of the *Drosophila* genome. Thus the number of HeLa proteins in each star category was multiplied by (1799/2827) to account for the smaller number of proteins identified in the S2 set and further multiplied by the ratio of estimated genes in *Drosophila* relative to humans (~14,000/21,000). For example, there are $93 \times 1799/2827 \times 14,000/21,000 = 39$ proteins classified as two- or three-star CCV components in the S2 set of proteins with complete profiles.

In Figure 2, all known CCV proteins predicted with two- or three-star confidence are annotated as CCV (in red), except clathrin, AP-1, and AP-2 subunits, which are shown in black, blue, and green, respectively. In Figure 3C, proteins predicted as CCV proteins with two- or three-star confidence are annotated as CCV (red) if their human homologues are known CCV proteins, except clathrin, AP-1, and AP-2 subunits, which are shown in black, blue, and green, respectively.

Abundance and stoichiometry estimation

The relative abundance of identified proteins was estimated from the mass spectrometric data using the iBAQ method (Schwanhauser *et al.*, 2011), as implemented in MaxQuant (Cox and Mann, 2008).

Each fractionation profiling data set consists of six sample pairs (SILAC heavy/light). In all cases, half the sample is the invariant “reference.” Thus for each protein there are up to six reference iBAQs. These were normalized to total reference iBAQ for each sample to account for differences in mass spectrometer performance between runs. The median iBAQ value was then calculated for each protein. All median iBAQs were normalized to the iBAQ of clathrin heavy chain (one of the most abundant proteins in the preparation). Clathrin heavy chain was then arbitrarily assigned an “abundance score” of 1,000,000. Hence a protein with an abundance score of 10,000 is present with 1% of the copy number of clathrin.

For the stoichiometry estimates in Figures 4 and 5 and Supplemental Figure S1, abundance scores were additionally normalized to the median abundance score of the core subunits of the complex of interest. Figures show the fold differences relative to the median score, on a \log_2 scale, to create equal distances for substoichiometric and superstoichiometric deviations. As a robust measure of the spread of individual abundance estimations, the median absolute deviation from the median (MAD) abundance score was calculated for each protein (indicated in Supplemental Tables S1 and S3). Error bars in Figures 4 and 5 and Supplemental Figure S1 also show MADs.

Cell culture and SILAC labeling

HeLa cells were maintained and SILAC (Ong *et al.*, 2002) labeled as described (Borner *et al.*, 2012). *Drosophila* S2 cells were grown in Schneider’s medium supplemented with 10% (vol/vol) fetal calf serum at 26–28°C in 175-cm² tissue culture flasks. SILAC labeling was performed based on the method described in Bonaldi *et al.* (2008). Schneider’s medium for SILAC labeling (without arginine and lysine) was purchased from Dundee Cell Products (Dundee, UK).

Gel electrophoresis, Western blotting, and immunofluorescence microscopy

Gels were Coomassie stained and scanned as described (Antrobus and Borner, 2011). Western blotting was performed as in Borner *et al.* (2012). Immunofluorescence microscopy was performed as described for Dmel2 cells (Hirst *et al.*, 2009). The following antibodies were used: dAP1G (Hirst *et al.*, 2009), dCLC (Heerssen *et al.*, 2008), and anti-V5 (46-0705; Life Technologies). Images were adjusted for brightness and contrast in Photoshop (Adobe Systems Europe, Maidenhead, UK) or PowerPoint (Microsoft).

Drosophila constructs and transient gene expression

LqfR-V5, SCYL2-V5, SMAP2-V5, and SES1/2-V5 constructs for expression in S2 cells were assembled in the constitutive vector pAc5.1V5 (Life Technologies). Genes were amplified using RedTaq from *Drosophila* Dmel2 cDNA made in-house from Dmel2 lysates, and various restriction sites were added at the 5’ and 3’ ends: SCYL2 (*Eco*R1 and *Apa*1), LqfR (*Kpn*1 and *Apa*1), SES1/2 (*Eco*R1 and *Xba*1), and SMAP2 (*Eco*R1 and *Xba*1). All constructs were fully sequenced; SMAP2 and SES1/2 were free of all changes, and LqfR and SCYL2 had one conservative substitution each. Constructs were transfected using TransIT-2020 (Mirus Bio, Madison, WI) following the manufacturer’s instructions, and cells were fixed 5–7 d posttransfection to moderate expression levels. In brief, for each transfection, we combined 1 μ l of TransIT and 500 μ g of DNA in 50 μ l of serum-free medium, added this to 0.6×10^6 cells in serum-free medium, incubated for 4 h, and then recovered to full growth medium.

Electron microscopy

Pelleted fractions were fixed with 2% paraformaldehyde/2.5% glutaraldehyde in 0.1 M cacodylate buffer, pH 7.2, at room tempera-

ture for 30 min, washed with 0.1 M cacodylate buffer, and postfixed using 1% osmium tetroxide for 1 h. Pellets were then washed before being incubated with 1% tannic acid in 0.05 M cacodylate buffer, pH 7.2, for 40 min to enhance contrast. Pellets were dehydrated in ethanol before being embedded in Araldite CY212 epoxy resin (Agar Scientific, Stansted, UK). Ultrathin sections (70 nm) were cut using a diamond knife mounted to a Reichert Ultracut S ultramicrotome (Leica Microsystems, Milton Keynes, UK) and picked up onto coated electron microscopy grids. The sections were stained with lead citrate and observed in a FEI Tecnai Spirit (Eindhoven, Netherlands) transmission electron microscope at an operating voltage of 80 kV.

Fractionation profiling: conceptual aspects, limitations of the method, and suggestions for future applications

We expect that fractionation profiling in its present form can be used for the characterization of clathrin-coated vesicles from most cell types, provided they are amenable to SILAC labeling and mechanical lysis. In addition, the approach can be adapted to other biological questions. Here we discuss key aspects of the method that we consider critical to the success of future applications.

Suitable targets

Fractionation profiling is best suited for the analysis of vesicles of fairly uniform size and density, as these produce sharply defined abundance distribution profiles. We expect the method to perform well with different types of coated vesicles, synaptic vesicles, and dense-core vesicles. In addition, our data show that large protein particles, such as proteasomes, can also generate highly characteristic profiles. Conversely, if a vesicle population is very heterogeneous (e.g., intrinsically heterogeneous or rendered variable through fragmentation during cell lysis), it is likely to be more evenly distributed across the gradient, and such a profile may be less easily discerned from those of other heterogeneous vesicles. However, even in those cases, the method will generate useful information, albeit in a more limited way. Subunits of stable, obligatory protein complexes will always produce similar profiles, and these are usually very tightly clustered.

In essence, fractionation profiling is capable of providing insights at two cellular levels: whole vesicles and particles, as well as smaller assemblies, such as protein complexes and membrane domains. The difference is the “range” of the predicted associations. For homogeneous vesicles such as CCVs, it is possible to predict the complement of the entire structure, as well as to predict protein complexes within the assembly (e.g., the AP-1 and AP-2 complexes are readily predicted as individual complexes and as components of CCVs). For other proteins with less characteristic profiles, it is possible to predict immediately close proteins with confidence but not the composition of the associated compartment. An example is the AP-3 complex, which has a broad distribution in our analysis (Figure 1D). The different subunits are readily predicted as a tight complex, as is the known interaction with the BLOC-1 complex, but there are few high-confidence predictions beyond these immediate associations. This suggests that our vesicle preparation contains endosomal membrane patches coated with AP-3 domains but perhaps not many free AP-3 vesicles (which would most likely produce a sharper profile). Thus, if the aim of a fractionation analysis is to characterize a whole organelle, a fairly homogeneous population will be required. The analysis of tight protein complexes, however, will work even with very heterogeneous preparations.

Profile optimization

The most discriminating profiles show strong differences among the different subfractions, ideally with a clear peak or depletion in one

fraction. The CCV profile is a case in point, with a moderate peak in fraction 1 and strong depletion from fraction 3. Of importance, the depletion from fraction 3 is by far the most characteristic trait of the CCV profile and contributes most to setting it apart from other profiles.

Spin conditions (speed and duration) will need to be optimized for each individual target vesicle (initially by Western blotting for marker proteins). The speeds we used here for CCVs may serve as a starting point, as most membrane vesicles will pellet in this range. Ideally, the different subfractions should have fairly similar total protein content to simplify downstream processing.

Note that vesicle profiles may differ between cell types. For example, the profiles of CCVs from HeLa and S2 are similar but not identical. Analysis of several marker proteins associated with the organelle of interest (e.g., by Western blotting) may help to identify such profile shifts and, if required, fine-tune the spin conditions. In the case of S2 cells, however, cell-type-specific optimization was not required; small shifts do not appear to pose a problem.

Sample purity

Every fractionation profiling experiment begins with a preparation enriched in the organelle of interest. How enriched/pure this preparation needs to be is difficult to predict in general. As a reference, we quantified the proportion of CCV proteins in our preparation, based on the relative abundance estimates that are part of the Predictor output. Briefly, we summed the relative abundance scores of all CCV proteins and divided this by the summed abundance scores of all proteins in the preparation. This gives the proportion of protein molecules (copy number) associated with CCVs in the preparation. Next we weighted these numbers by the molecular weights of individual proteins, to estimate the proportion of protein mass associated with CCVs. According to these calculations, the HeLa CCV preparation contains 4.6% CCV proteins (proportion of copy number) and ~9.6% CCV protein (proportion of total protein mass). The main reason for the discrepancy between these numbers is that clathrin heavy chain—the predominant CCV protein—has a very high molecular weight. The S2 preparation is about half as pure, with 2.4% CCV proteins (copy number) and 5.1% CCV protein (total mass).

Although we have not investigated the requirements for CCV enrichment systematically, it seems likely that higher purity will improve the analysis. First, it will increase the number of identified relevant proteins. It will also increase the average number of quantification events per protein and thus the quantification accuracy. Of importance, if a protein is associated with more than one organelle, it will only be classified as a CCV protein if its predominant association within the preparation is with CCVs, and so a high proportion of CCVs in the preparation will enhance the discriminating power (see also the final section, *Caveats and limitations of the approach*). Nevertheless, as our analysis shows, even a preparation that contains >95% “contaminating” non-CCV proteins is readily amenable to fractionation profiling.

Number of subfractions

The number of analyzed subfractions is another important aspect of the experimental design. Here we tried to minimize this number to keep the mass spectrometric analysis time as short as possible while maintaining high accuracy of prediction for CCVs. Theoretically, a larger number of fractions or experimental repeats will improve accuracy, but the trade-off is that more gaps will appear in the data matrix, as proteins are less likely to be identified in all subfractions. In addition, the yield of the CCV preparation is rather low, and if this

low amount of protein is distributed over more subfractions, it may decrease the number of protein identifications and indirectly (via reduced sequencing coverage) even the quantification accuracy. Finding the optimum number of subfractions hence depends on a number of considerations. The more complex the mixture, the more fractions are advisable, if the quantity of starting material is not limiting. For CCVs, two biological replicates of three fractions (i.e., six data points) yield very solid data; for other organelles or subcellular preparations, the number will need to be optimized case by case. Based on our experience, we expect that 6–10 data points will be sufficient for most applications.

Mass spectrometric analysis

Our experimental setup maximizes the accuracy of the quantitative mass spectrometry. Tight profile clustering is hence achieved with a relatively small number of subfractions/data points. Key to this strategy is the use of the SILAC method (Ong *et al.*, 2002), which is extremely accurate for measuring small or medium differences in protein abundance by mass spectrometry. The approach is based on metabolic labeling of cells in culture by supplementing the growth medium with different isotopes of the amino acids arginine and lysine before subcellular fractionation. All proteins from the “reference” sample are distinguishable from proteins in the subfractions by a mass shift. This allows pooling of reference and subfraction samples before enzymatic digestion for mass spectrometric analysis, thus avoiding variability introduced by sample preparation. Furthermore, mass spectrometric analysis of SILAC samples compares ion intensities of isotopic variants of the same peptide within the same mass spectrometry run. This avoids normalization errors required for “label-free” quantification, which compares ion intensities of the same peptide between different mass spectrometry runs. Instrument performance differences between runs are therefore less problematic with SILAC. A drawback of the method is that the dynamic range of quantification is limited by the isotope labeling efficiency. In practice, SILAC is extremely accurate for the determination of relative differences less than ~20-fold. In fractionation profiling, the differences in protein abundance between subfractions and the reference fraction are deliberately kept small to be within that range. Furthermore, the subfractions and the reference have almost the same overall protein composition; they differ mostly in the relative abundance of proteins (i.e., the same species of proteins are present but in different proportions). This facilitates the identification of proteins across all subfractions, avoiding gaps in the data matrix. Finally, the high-performance mass spectrometer (Q Exactive) we used here allows the quantification of multiple peptides/SILAC pairs per fraction for each protein, resulting in very robust quantification.

Caveats and limitations of the approach

Fractionation profiling identifies groups of proteins with similar fractionation behavior and, by inference, similar subcellular distribution at steady state. Both false-positive and -negative classifications can occur. Similar profiles may arise by chance; structures with similar size and density will have similar fractionation properties and thus may not be resolved by the method. The resolution is also lower for profiles near the baseline (i.e., broad uniform distribution across the subfractions), and the Predictor alerts users when queried with such a profile.

Conversely, there are several reasons why the predictor may not show a previously established interaction. First, the interaction may be transient and thus not substantial at steady state. Second, a binding partner may be mostly engaged in another interaction and thus

profile with this predominating partner. Third, a protein may have multiple stable binding partners, resulting in a “mixed” profile.

A related situation arises when a protein is present at multiple subcellular localizations. As with multiple binding partners, the outcome depends highly on the composition of the preparation. For example, assume that protein X is equally found in compartments A and B. If A and B are both present in similar proportions in the sample preparation, protein X will have a hybrid profile, which may be difficult to interpret or even without predictive value. If compartment A is strongly enriched in the preparation but compartment B is not, then the A profile will dominate, and vice versa. A good example of this situation is the cation-independent mannose 6-phosphate receptor (IGF2R), a cargo protein of CCVs. It is abundant in CCVs but also present at the *trans*-Golgi network and on endosomes, fragments of which are present in the vesicle fraction. Because the CCVs are strongly enriched in the preparation compared with the other membrane compartments, IGF2R has a clear CCV profile. This example further highlights how fractionation profiling benefits from a sample preparation highly enriched in the organelle of interest.

ACKNOWLEDGMENTS

We thank Paul Luzio, Sebastian Schuck, and members of the Robinson and Mann groups for their critical feedback. G.H.H.B., J.H., J.R.E., and M.S.R. were funded by the Wellcome Trust (Grant RG52996). In addition, G.H.H.B. received an EMBO Short Term Fellowship, which made this collaboration possible. M.Y.H. and M.M. were supported within a project framework of German medical genome research funded by the Bundesministerium für Bildung und Forschung through Grant FKZ01GS0861 in the DiGtoP consortium.

REFERENCES

- Allen CL, Goulding D, Field MC (2003). Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J* 22, 4991–5002.
- Antrobus R, Borner GH (2011). Improved elution conditions for native co-immunoprecipitation. *PLoS One* 6, e18218.
- Bassik MC, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, Weibezahn J, Horlbeck MA, Chen S, Mann M, et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922.
- Blondeau F, Ritter B, Allaire PD, Wasiak S, Girard M, Hussain NK, Angers A, Legendre-Guillemain V, Roy L, Boismenu D, et al. (2004). Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc Natl Acad Sci USA* 101, 3833–3838.
- Bonaldi T, Straub T, Cox J, Kumar C, Becker PB, Mann M (2008). Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol Cell* 31, 762–772.
- Bonnemaison ML, Eipper BA, Mains RE (2013). Role of adaptor proteins in secretory granule biogenesis and maturation. *Front Endocrinol* 4, 101.
- Borner GH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, Sahlender DA, Robinson MS (2012). Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J Cell Biol* 197, 141–160.
- Borner GH, Harbour M, Hester S, Lilley KS, Robinson MS (2006). Comparative proteomics of clathrin-coated vesicles. *J Cell Biol* 175, 571–578.
- Borner GH, Rana AA, Forster R, Harbour M, Smith JC, Robinson MS (2007). CVAK104 is a novel regulator of clathrin-mediated SNARE sorting. *Traffic* 8, 893–903.
- Cox J, Mann M (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367–1372.
- Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10, 1794–1805.
- Dunkley TP, Watson R, Griffin JL, Dupree P, Lilley KS (2004). Localization of organelle proteins by isotope tagging (LOPIT). *Mol Cell Proteomics* 3, 1128–1134.
- Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O’Connor L, Li M, et al. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3, 89.
- Eyster CA, Higginson JD, Huebner R, Porat-Shliom N, Weigert R, Wu WW, Shen RF, Donaldson JG (2009). Discovery of new cargo proteins that enter cells through clathrin-independent endocytosis. *Traffic* 10, 590–599.
- Faini M, Beck R, Wieland FT, Briggs JA (2013). Vesicle coats: structure, function, and general principles of assembly. *Trends Cell Biol* 23, 279–288.
- Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M (2006). A mammalian organelle map by protein correlation profiling. *Cell* 125, 187–199.
- Girard M, Allaire PD, McPherson PS, Blondeau F (2005). Non-stoichiometric relationship between clathrin heavy and light chains revealed by quantitative comparative proteomics of clathrin-coated vesicles from brain and liver. *Mol Cell Proteomics* 4, 1145–1154.
- Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081.
- Heerssen H, Fetter RD, Davis GW (2008). Clathrin dependence of synaptic-vesicle formation at the *Drosophila* neuromuscular junction. *Curr Biol* 18, 401–409.
- Heisler FF, Loebrich S, Pechmann Y, Maier N, Zivkovic AR, Tokito M, Hausrat TJ, Schweizer M, Bähring R, Holzbaur EL, et al. (2011). Muskelein regulates actin filament- and microtubule-based GABA(A) receptor transport in neurons. *Neuron* 70, 66–81.
- Hirst J, Borner GH, Antrobus R, Peden AA, Hodson NA, Sahlender DA, Robinson MS (2012). Distinct and overlapping roles for AP-1 and GGAs revealed by the “knocksideways” system. *Curr Biol* 22, 1711–1716.
- Hirst J, Sahlender DA, Choma M, Sinka R, Harbour ME, Parkinson M, Robinson MS (2009). Spatial and functional relationship of GGAs and AP-1 in *Drosophila* and HeLa cells. *Traffic* 10, 1696–1710.
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O’Shea EK (2003). Global analysis of protein localization in budding yeast. *Nature* 425, 686–691.
- Kobayashi N, Yang J, Ueda A, Suzuki T, Tomaru K, Takeno M, Okuda K, Ishigatsubo Y (2007). RanBPM, Muskelein, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* 396, 236–247.
- Kristensen AR, Gsponer J, Foster LJ (2012). A high-throughput approach for measuring temporal changes in the interactome. *Nat Methods* 9, 907–909.
- Leto D, Saltiel AR (2012). Regulation of glucose transport by insulin: traffic control of GLUT4. *Nature reviews. Mol Cell Biol* 13, 383–396.
- Menssen R, Schweiggert J, Schreiner J, Kusevic D, Reuther J, Braun B, Wolf DH (2012). Exploring the topology of the Gid complex, the E3 ubiquitin ligase involved in catabolite-induced degradation of gluconeogenic enzymes. *J Biol Chem* 287, 25602–25614.
- Montpetit A, Cote S, Brustein E, Drouin CA, Lapointe L, Boudreau M, Meloche C, Drouin R, Hudson TJ, Cossette P (2008). Disruption of AP1S1, causing a novel neurocutaneous syndrome, perturbs development of the skin and spinal cord. *PLoS Genet* 4, e1000296.
- Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol Cell Proteomics* 11, M111.013722.
- Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 7, 548.
- Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376–386.
- Pierce BL, Kibriya MG, Tong L, Jasmine F, Argos M, Roy S, Paul-Brutus R, Rahaman R, Rakibuz-Zaman M, Parvez F, et al. (2012). Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet* 8, e1002522.
- Pinheiro C, Longatto-Filho A, Azevedo-Silva J, Casal M, Schmitt FC, Baltazar F (2012). Role of monocarboxylate transporters in human cancers: state of the art. *J Bioenerg Biomembr* 44, 127–139.
- Rappsilber J, Mann M, Ishihama Y (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2, 1896–1906.
- Robinson MS (2004). Adaptable adaptors for coated vesicles. *Trends Cell Biol* 14, 167–174.

- Romani B, Cohen EA (2012). Lentivirus Vpr and Vpx accessory proteins usurp the cullin4-DDB1 (DCAF1) E3 ubiquitin ligase. *Curr Opin Virol* 2, 755–763.
- Ruepp A, Waegelé B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res* 38, D497–501.
- Sandvig K, Pust S, Skotland T, van Deurs B (2011). Clathrin-independent endocytosis: mechanisms and function. *Curr Opin Cell Biol* 23, 413–420.
- Schwanhauser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1, 2856–2860.
- Tarpey PS, Stevens C, Teague J, Edkins S, O’Meara S, Avis T, Barthorpe S, Buck G, Butler A, Cole J, et al. (2006). Mutations in the gene encoding the Sigma 2 subunit of the adaptor protein 1 complex, AP1S2, cause X-linked mental retardation. *Am J Hum Genet* 79, 1119–1124.
- Tomko RJ Jr, Hochstrasser M (2013). Molecular architecture and assembly of the eukaryotic proteasome. *Annu Rev Biochem* 82, 415–445.
- Umasankar PK, Sanker S, Thieman JR, Chakraborty S, Wendland B, Tsang M, Traub LM (2012). Distinct and separable activities of the endocytic clathrin-coat components Fcho1/2 and AP-2 in developmental patterning. *Nat Cell Biol* 14, 488–501.
- Valiyaveetil M, Bentley AA, Gursahany P, Hussien R, Chakravarti R, Kureishy N, Prag S, Adams JC (2008). Novel role of the muskellin-RanBP9 complex as a nucleocytoplasmic mediator of cell morphology regulation. *J Cell Biol* 182, 727–739.
- van Weering JR, Verkade P, Cullen PJ (2010). SNX-BAR proteins in phosphoinositide-mediated, tubular-based endosomal sorting. *Semin Cell Dev Biol* 21, 371–380.
- Yebenes H, Mesa P, Munoz IG, Montoya G, Valpuesta JM (2011). Chaperonins: two rings for folding. *Biochem Sci* 36, 424–432.
- Yu S, Liang Y (2012). A trapper keeper for TRAPP, its structures and functions. *Cell Mol Life Sci* 69, 3933–3944.