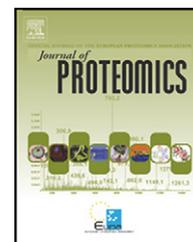


available at www.sciencedirect.comwww.elsevier.com/locate/jprot

Review

When peptides fly: Advances in *Drosophila* proteomics

Alexey Veraksa*

Department of Biology, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125, USA

ARTICLE INFO

Keywords:

Drosophila
Proteomics
Mass spectrometry
Affinity purification
Proteome
Model organism

ABSTRACT

In the past decade, improvements in genome annotation, protein fractionation methods and mass spectrometry instrumentation resulted in rapid growth of *Drosophila* proteomics. This review presents the current status of proteomics research in the fly. Areas that have seen major advances in recent years include efforts to map and catalog the *Drosophila* proteome and high-throughput as well as targeted studies to analyze protein–protein interactions and post-translational modifications. Stable isotope labeling of flies and other applications of quantitative proteomics have opened up new possibilities for functional analyses. It is clear that proteomics is becoming an indispensable tool in *Drosophila* systems biology research that adds a unique dimension to studying gene function.

© 2010 Elsevier B.V. All rights reserved.

Contents

1. Introduction	2158
2. Advantages of using flies as a platform for proteomics research	2159
3. Mapping and cataloging the fly proteome	2159
4. Analysis of protein–protein interactions in <i>Drosophila</i> : the yeast two hybrid (Y2H) approaches	2161
5. Analyzing protein complexes by affinity purification-mass spectrometry (AP-MS)	2161
6. Mapping post-translational modifications (PTMs) in fly proteins	2163
7. Revealing proteome dynamics: quantitative proteomics in <i>Drosophila</i>	2165
8. Summary and outlook	2166
Acknowledgements	2167
References	2167

1. Introduction

In the past two decades, high-throughput approaches to studying cellular functions have moved from the analysis of genomes and transcriptomes to the realm of proteins. Proteomics is now a rapidly evolving field of research, thanks to such advances as mass spectrometry identification of proteins, improvements in

methods to analyze protein–protein interactions, and development of high-throughput approaches to synthesize and purify proteins for microarray and crystallography applications. Proteomics is an integral part of systems-level analysis of cellular functions, promising to shed light on regulatory mechanisms that are beyond the reach of genomics and transcriptomics [1].

The fruit fly *Drosophila melanogaster* has been a workhorse model organism for over 100 years, first as a favorite system to study fundamental genetic principles, and more recently as an experimentally tractable organism to analyze higher eukaryotic

* Tel.: +1 617 287 6665; fax: +1 617 287 6650.
E-mail address: alexey.veraksa@umb.edu.

genomes and to model human diseases [2,3]. The past 10 years witnessed the birth and rise of *Drosophila* as a model system for proteomics studies. Examination of PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) references shows that by the end of 2009 articles on *Drosophila* proteomics comprised approximately 1.1% of all published papers on proteomics, which may seem at a first glance like a low number (Fig. 1A). However, plotting cumulative percentages of the respective groups of articles shows that *Drosophila* proteomics is growing at a rate that is very close to that of the general field of proteomics (Fig. 1B). In fact, proteomics in the fly compares favorably to the whole field of *Drosophila* research which constitutes approximately 0.5% of all papers in PubMed.

Efforts to characterize the *Drosophila* proteome that have been carried out to date can be grouped into the following areas of investigation: i) mapping and cataloging the proteome, ii) analysis of protein–protein interactions, iii) analysis of post-translational protein modifications; iv) quantitative analysis of the proteome, and v) functional studies that often combine more than one approach. In this review, I describe how these studies contribute to our knowledge of the organization and function of the *Drosophila* proteome. I also summarize challenges and future directions for *Drosophila* proteomics.

2. Advantages of using flies as a platform for proteomics research

Protein identification by mass spectrometry has become a primary method of analysis in proteomics [1]. This method relies on genome annotation to provide accurate gene models that correctly predict protein coding regions. The *Drosophila* genome [4] is arguably one of the best annotated, thanks to extensive curation and annotation carried out at FlyBase, a central repository of genomic and proteomic information on

Drosophila (<http://flybase.org/>). In addition to a well-annotated genome, *Drosophila* offers a unique possibility to rapidly move from high-throughput screening to functional assays, because mutant alleles and transgenic constructs, including inducible RNAi reagents, are available for most genes. Knowledge of cellular mechanisms obtained in flies is in many cases applicable to human biology, because many of the regulatory mechanisms and signaling pathways are conserved between flies and mammals. This property makes *Drosophila* an attractive model system for several human diseases [3,5].

Proteomics studies often require significant amounts of starting material that is subjected to various types of fractionation, and fortunately, *Drosophila* development lends itself to large-scale biochemical experiments. The life cycle includes four distinct developmental stages: embryo (first 24 h of development after egg laying), three larval stages separated by molts (combined duration is 4 days), pupal stage during which metamorphosis occurs (4–4.5 days), and an adult that under normal laboratory conditions lives for about a month [6]. These different developmental stages of the life cycle can be easily separated and collected in large quantities as synchronized populations of individuals. As an additional benefit, availability of widely used cultured cell lines (such as Schneider's S2 and Kc167 lines) facilitates proteomic analyses in *Drosophila*. These and other cell lines are available from the *Drosophila* Genomics Resource Center (<https://dgrc.cgb.indiana.edu/>).

3. Mapping and cataloging the fly proteome

Some statistics on the *D. melanogaster* genome and proteome are summarized in Table 1. The fly genome is rather compact at 1.8×10^8 bases, which is an order of magnitude smaller than the mammalian genome [4]. A recent FlyBase annotation (release 5.25) lists 13,781 protein coding genes (gene models), which

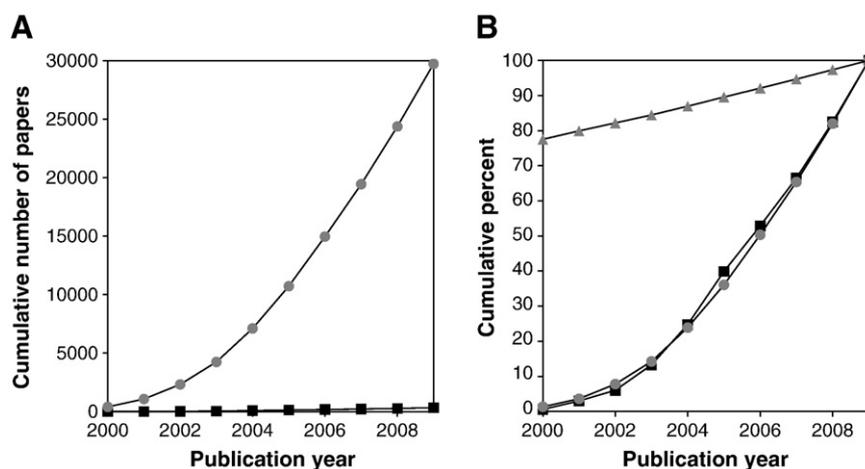


Fig. 1 – Growth of *Drosophila* proteomics: a view from PubMed. (A) Most of the papers on proteomics were published in the past 10 years. Cumulative numbers of papers as referenced in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>), published by and including the year shown, ending with 2009. Gray circles, papers found using the search term “proteomic OR proteomics”. Black squares, papers found using the search term “*Drosophila* AND (proteomic OR proteomics)”. (B) The field of *Drosophila* proteomics is growing at the same rate as the whole field of proteomic research. Cumulative percent of all papers published by and including the year shown. The number of papers published through 2009 was set as 100%. Gray circles and black squares indicate same searches as in (A). Growth of a mature field is shown for comparison (gray triangles, papers found using the search term “ribosome”).

Table 1 – Statistics of the *Drosophila* genome and proteome.**Source: Berkeley *Drosophila* Genome Project genome release 5, FlyBase annotations release 5.25.**

Sequenced euchromatic genome	120,381,546 bp
Sequenced heterochromatic genome	19,330,818 bp
Protein coding genes (gene models)	13,781
Protein coding transcripts	21,909
Unique polypeptides	18,358
Average polypeptide length	601 aa
Longest polypeptide	22,971 aa
Shortest polypeptide	11 aa

aa, amino acid.

are predicted to encode 18,358 unique protein isoforms. The latter number is likely an underestimate due to our limited knowledge of a full repertoire of splice variants for most proteins. Systematic mapping of the proteome content presents a considerable challenge, since expression levels of proteins can vary over several orders of magnitude, and physiological protein samples often have a bewildering complexity [7]. Introduction of mass spectrometry based protein identification methods has made high-throughput analysis of proteomes a reality, yet no single proteome has been completely mapped [1].

Some of the efforts to catalog the proteome content in *Drosophila* have focused on the analysis of individual tissues or organelles (Fig. 2). One of the first attempts to comprehensively characterize the *Drosophila* proteome employed 2-dimensional gel electrophoresis and established a reference protein map of the larval wing imaginal discs [8]. Since that report, multiple studies characterized and cataloged the proteome subsets in

different tissues, such as adult heads and embryos [9,10], male reproductive system [11], sperm [12,13] (also see a review by T. Karr in this issue), wing imaginal discs [14], and larval hemolymph [15–21]. Other studies carried out the analysis of whole secreted neuropeptides, hormones, and peptide precursors by the so-called “peptidomics” approaches [22–27]. At the subcellular level, researchers analyzed the protein composition of the *Drosophila* mitochondria [28], ribosomes [29], and lipid droplets from the larval fat bodies [30,31].

These focused studies provided representations of the proteome content in specific organs, tissues, and cellular organelles. Improvements in the sensitivity and resolution of mass spectrometers made it possible to carry out large-scale systematic surveys with the goal of identifying every predicted protein in the *Drosophila* proteome. In a seminal study [7], R. Aebersold et al. described a high-quality catalog of the *D. melanogaster* proteome covering 9124 distinct proteins, or 63% of the protein-encoding gene models, at 1.37% false discovery rate. This level of proteome coverage has not been achieved for any other multicellular eukaryote, and was made possible by a combination of approaches to maximize sample diversity through analysis of different cell types and developmental stages, multidimensional biochemical fractionation to reduce sample complexity, and analysis-driven experimental feedback strategy in which statistical analysis of prior data guided the design of the next set of proteomic experiments [7,32]. The data from this study is publicly available through PeptideAtlas (<http://www.peptideatlas.org>) and is continuously updated. A recent release increased the coverage to 9263 protein isoforms corresponding to 8799 (65%) gene models [33]. There are several reasons why proteins corresponding to the remaining 35% of

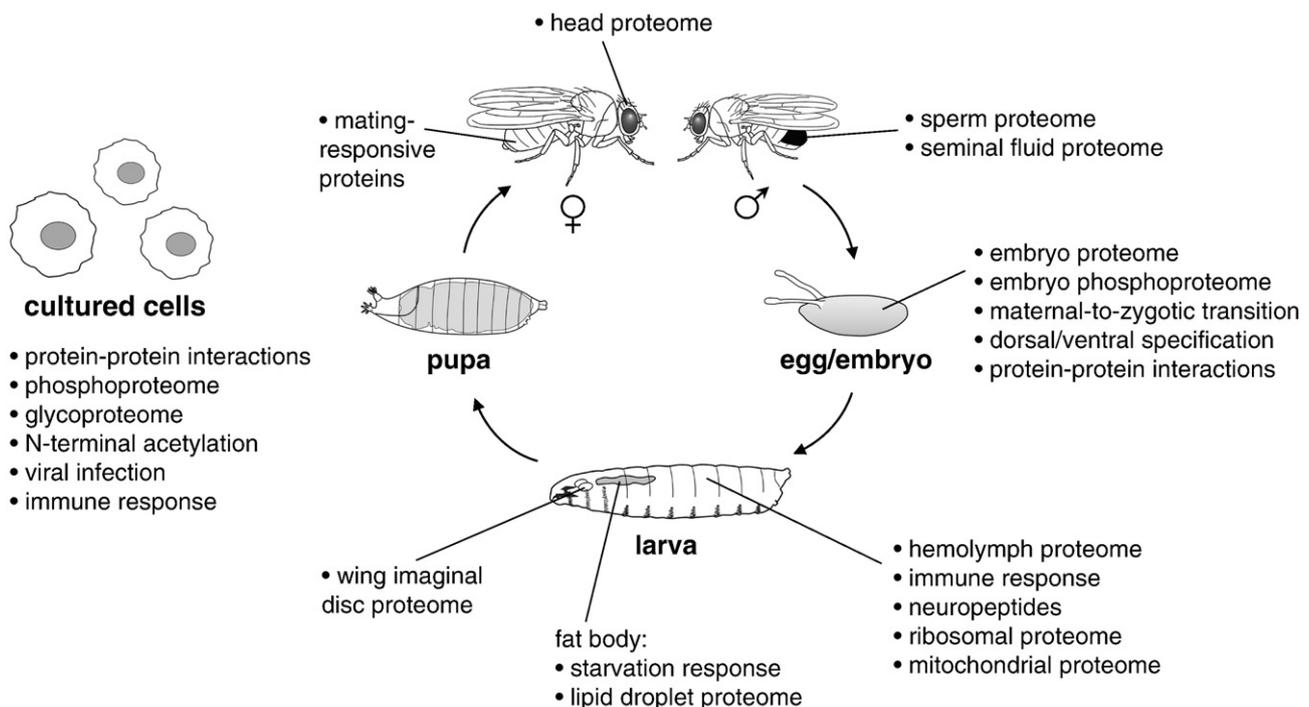


Fig. 2 – Biological processes and subsets of the *Drosophila* proteome that were a subject of proteomics studies. Left, experiments in cultured cells such as the S2 and Kc167 cell lines. Right, stages of the fly life cycle, with indicated areas of proteomics studies. See text for description of individual experiments and the corresponding references.

gene models were not observed. These proteins that remain to be discovered may include polypeptides that are unstable, expressed at low levels, or those that require further fractionation [33]. The hope is that additional extraction, separation and enrichment methods will further improve the coverage of the proteome.

How does the data from PeptideAtlas advance proteomic experiments in *Drosophila*? First, it provides a resource for new shotgun proteomics studies that can now employ spectral library searching rather than traditional searches of protein databases. In a proof of principle experiment, it was shown that the sensitivity, specificity, and speed of peptide identification can be improved by searching pre-acquired spectra that are processed into consensus spectra [33]. However, it would seem that proteome coverage would have to be significantly improved in order for this method to become a truly general protein identification approach. Second, PeptideAtlas is a resource for targeted proteomics experiments that make use of proteotypic peptides (PTPs), which are peptides that can unambiguously identify a protein of interest in a complex mixture. The use of PTPs allows for absolute protein quantification in selected reaction monitoring experiments (SRM, also known as multiple reaction monitoring, MRM) performed on triple quadrupole instruments [7,32]. Third, a high-quality proteome catalog can be used for improving genome annotation, for example by validating features of protein coding gene models such as splicing isoforms, and even discovering previously unannotated genes. At present, the information from PeptideAtlas is being integrated with FlyBase and is now visible as an option in the “Genome reagents and data” track in the GBrowse interface. Finally, large-scale proteome mapping facilitates comparative studies. Thus, a recent deep analysis of the *Caenorhabditis elegans* proteome has revealed a surprisingly strong overall correlation of protein abundances between *C. elegans* and *Drosophila* (Spearman rank correlation $R_s=0.79$), which was much higher than the correlation of the corresponding transcript levels between the two species ($R_s<0.5$) [34]. In sum, the *Drosophila* PeptideAtlas and the associated software suite form a valuable resource for analyzing the proteome and designing new proteomic experiments.

4. Analysis of protein–protein interactions in *Drosophila*: the yeast two hybrid (Y2H) approaches

The knowledge of protein–protein interactions (PPIs) is an integral component of systems biology analysis, as cellular processes are carried out by networks of interacting proteins. The utility of studying PPIs is already evident even from the existing incomplete maps. They provide insights into the organization of biological networks, assist in determining functions of many proteins and protein complexes, and identify connections to proteins associated with human diseases [35].

While a complete coverage of the proteome remains a difficult task, mapping of PPIs presents an even bigger challenge. Two primary approaches have been utilized to study PPIs in a high-throughput format: yeast two hybrid analysis (Y2H) and affinity purification-mass spectrometry (AP-MS) (reviewed in [36–38]). The largest effort to date to map *Drosophila* PPIs by Y2H was reported in [39] (the so-called CuraGen dataset). This project

resulted in a construction of a draft map of 7048 proteins involved in 20,405 interactions, which corresponds to ~52% coverage of predicted gene models. An advantage of the Y2H approach is that it allows for a numerical estimation of the confidence of each PPI observation. Application of more stringent criteria to the Giot et al. data resulted in a high confidence map of 4679 proteins and 4780 interactions. It was estimated that 40% of the interactions in the high confidence map are likely to be biologically relevant [39]. It should be noted that the estimation of specificity and sensitivity of PPI maps is still a matter of debate, with different studies utilizing somewhat different approaches for obtaining such estimates [40]. Statistical analysis of the network properties showed that the *Drosophila* PPI network is best described by a two-level model: local connectivity likely representing interactions occurring within multiprotein complexes, and more global connectivity that potentially represents higher-order communication between complexes [39].

Additional large-scale Y2H projects have expanded the *Drosophila* binary PPI map. The Hybrigenics study used 102 bait proteins and detected over 2300 interactions, of which 710 were of high confidence [41]. Comparison with the CuraGen map found surprisingly little overlap between the two datasets. Thus, 30 bait proteins were common to both studies and yielded 216 PPIs in the CuraGen dataset and 662 PPIs in the Hybrigenics dataset. Unexpectedly, these two sets shared only 24 interactions [41]. This observation suggests that the two approaches yielded complementary information on the PPIs in *Drosophila* [41]. The Finley laboratory is continuously adding new interactions from Y2H screens, and is currently reporting 3161 interactions involving 1338 proteins. These data were obtained in a focused screen with 152 proteins related to cell cycle regulators [42] and a study that experimentally tested computationally predicted PPIs [43]. The Finley lab maintains a user-friendly repository of *Drosophila* PPIs, the *Drosophila* Interactions Database (DroID), with an updatable confidence score assigned to each interaction (<http://www.droidb.org/>, [44]). *Drosophila* PPIs can also be accessed via centralized databases such as the BioGRID (<http://www.thebiogrid.org/>).

Recently, a rigorous statistical approach for estimating a total number of binary PPIs in an organism’s proteome has been proposed [35,40,45]. Extrapolating from *C. elegans* estimates, a complete *Drosophila* binary protein interaction map is expected to contain approximately 10^5 pair-wise PPIs. This number makes it clear that our current coverage of the PPI space is far from complete, and that new, more efficient strategies for mapping PPIs have to be developed, in order to reduce analysis costs while maximizing the recovery and validation of new interactions. The study by Schwartz et al. provides a promising strategy for lowering the cost of completing the PPI maps, in which sample pooling is combined with prioritized testing and interaction prediction [43]. The authors point out that due to a high false negative rate of Y2H approaches (50–80%), multiple independent assay types will likely be needed to achieve complete coverage.

5. Analyzing protein complexes by affinity purification-mass spectrometry (AP-MS)

The AP-MS approach has emerged as a method that is largely complementary to Y2H [37,38]. While purification of whole

protein complexes does not give information about the binary interactions within the complex, it benefits from the fact that the complexes are formed in the natural cellular environment that provides relevant post-translational modifications and promotes correct protein folding. Also, some of the complexes are held together by cooperative interactions of several non-identical subunits, which cannot be achieved when analyzing binary interactions. Information about the composition of protein complexes can help assign function to previously uncharacterized proteins, based on the “guilt by association” principle, although Y2H studies can also be used for this purpose. A recent large-scale analysis of protein complexes in yeast has revealed modularity in the organization of the proteome, in which stable invariant core modules can recruit more variable subunits that may be shared between different complexes [46]. There is no doubt that large-scale analysis of the *Drosophila* proteome by AP-MS would yield valuable information that would complement the current Y2H data. Yet no such project has been carried out to date, but as mentioned below, at least one high-throughput effort is under way.

The “affinity” step of AP-MS can be performed by using a specific antibody against a protein of interest. In such experiments, native protein complexes are isolated from cells or tissues, thus maximally approximating the endogenous cellular conditions. However, this is impractical for large-scale studies, as antibodies are available for only a fraction of all proteins. Therefore, medium- to high-throughput AP-MS

approaches all rely on the use of affinity tags [47,48]. A variety of affinity tags has been used in *Drosophila* proteomics research. Many studies have made use of single tags to isolate the protein of interest and characterize the associated cellular proteins by mass spectrometry. The single tags employed ranged from more “traditional” ones such as FLAG [49] and β -galactosidase [50], to more recently introduced biotin ligase recognition peptide (BLRP, also known as AviTag™) [51] and TagIt [52]. Lichty et al. [48] compared the performance of a panel of single tags using extracts from *E. coli*, *Drosophila* and human cells, and concluded that the Strep II tag [53] may provide an acceptable compromise of excellent purification with good yields at a moderate cost [48].

Despite the demonstrated success of using single tags for protein complex purification, double tag combinations result in a significant reduction of contaminating components. Several variations of tandem affinity purification (TAP) have been proposed [54–57]. The most widely used combination has been the TAP tag developed by Rigaut et al., which consists of two Protein A modules and a calmodulin binding peptide (CBP), separated by a TEV protease cleavage site [58] (Fig. 3). The TAP tag has been successfully used both in large-scale studies and focused small-scale experiments in yeast and mammalian cells [46,59–62]. In the first study to implement the TAP approach in flies, human TAP-tagged proteins were expressed in *Drosophila* cultured cells and used as baits to recover the associated *Drosophila* proteins [63]. A concomitant RNAi-mediated knockdown of the *Drosophila*

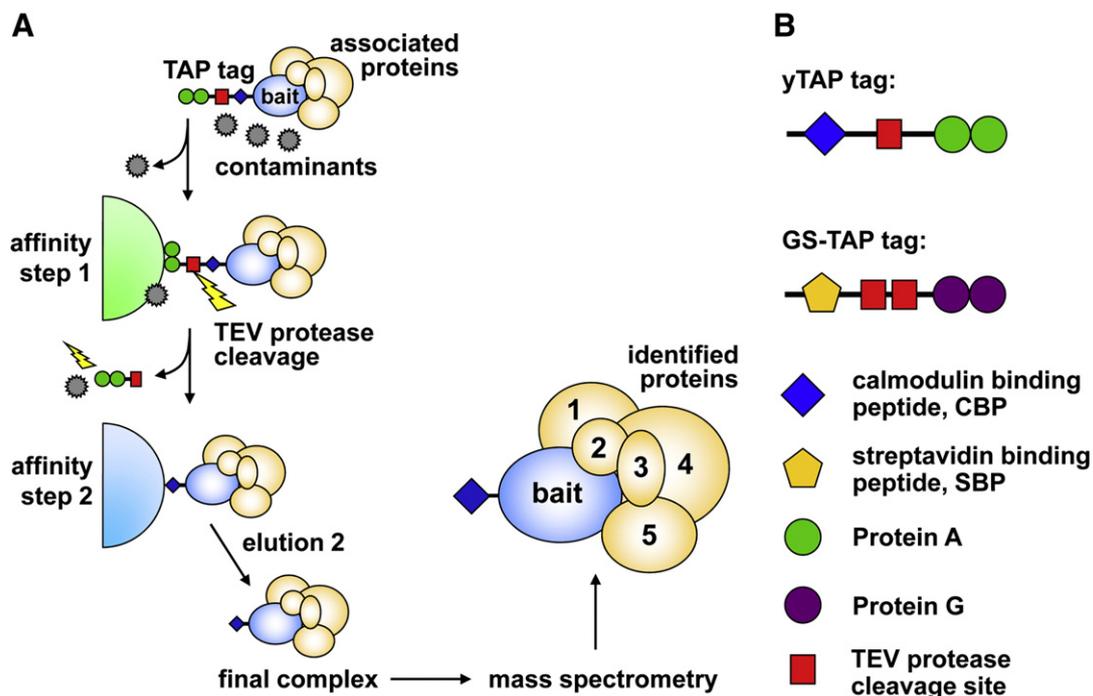


Fig. 3 – Outline of the tandem affinity purification (TAP) approach and the structure of the TAP tags. (A) A bait protein is tagged with the TAP tag and expressed in a tissue or cell line of interest. The bait and associated proteins are purified using two affinity steps. Cleavage with the TEV protease is used to release the protein complex after the first affinity step. Second elution yields a final complex which is subjected to protein identification by mass spectrometry and database searching. (B) Schematic diagram of the conventional TAP tag (yTAP tag) [58,64] and an improved version, the GS-TAP tag [75,76].

protein corresponding to the human bait improved the recovery of the interacting subunits. However, subsequent studies in mammalian cells that employed controlled overexpression of the TAP-tagged protein baits showed that biologically relevant protein complexes can be recovered without the need to reduce the level of the corresponding endogenous protein [62]. Ideally, however, the AP-MS approaches should use organisms in which the tagged protein is completely replacing its endogenous counterpart, as was done in the yeast studies, thus obviating potential problems associated with overexpression.

We have developed vectors for inducible expression of the TAP-tagged proteins in *Drosophila* cultured cells and in vivo, and used the TAP approach to analyze protein complexes in the Notch signaling pathway [64]. As with the results of any screening approach, a definitive answer about the validity of an observed interaction can only be obtained by functional analysis of the interacting partners in vivo. In subsequent studies, we used traditional genetic and biochemical approaches to validate the results of TAP experiments [65,66]. Other reports have used the TAP method and biologically validated the identified interactions [67–73], demonstrating that the TAP-MS approach yields meaningful insights into the organization of the PPI network in flies. More recently, we showed that TAP-MS also works in another favorite model organism, zebrafish [74].

The TAP technology was recently improved by using a new combination of tags. The GS-TAP tag developed by Burckstummer et al. consists of two Protein G modules and a streptavidin binding peptide (SBP), separated by one or two TEV protease cleavage sites [75] (Fig. 3B). The authors reported an up to 10-fold increase in the yield of the purified protein complexes using this new tag combination. We found that the GS-TAP tag preserves the function of the tagged proteins in *Drosophila* by using genetic rescue experiments [76]. While the GS-TAP procedure indeed results in higher yields of the purified protein complexes, a major advantage of using this system is in a dramatically reduced level of recovery of contaminating proteins, and a corresponding increase in signal-to-noise ratio for the bona fide protein complex components, compared to the original TAP tag [76]. The latter consideration is of particular importance for the subsequent mass spectrometry applications, where identification of low abundance interactors may be obstructed by high levels of contaminants. We found that most of the improvement was due to the use of the SBP tag [77] in the GS-TAP system rather than CBP. In pilot experiments under way in our laboratory, single-step streptavidin-based purifications using the SBP tag alone or even as part of the larger GS-TAP tag, followed by elution with biotin, have yielded protein complexes of sufficient purity for subsequent analysis by mass spectrometry, suggesting that the SBP tag can be used as a viable alternative to other widely used streptavidin-based systems such as the Strep II tag.

While high-throughput AP-MS data for *Drosophila* is conspicuously missing, at least two projects are being carried out to fill this gap in our knowledge of the fly proteome. A large-scale multi-laboratory collaborative project, led by S. Artavanis-Tsakonas and involving the S. Celniker, S. Gygi, and K. VijayRaghavan labs, is currently under way with the goal of

generating a comprehensive protein complex map of the *Drosophila* interactome. A primary approach used is tagging proteins with an HA tag and analyzing protein complexes by AP-MS after expression in S2R+ cells. The target bait space includes approximately 8000 full-length expression constructs generated by the Berkeley *Drosophila* Genome Project (<http://www.fruitfly.org>), which corresponds to ~58% of the gene models. In a parallel set of experiments, the same set of bait proteins will be expressed in vivo in transgenic flies, and the resulting protein complexes will be analyzed by AP-MS. Completion of this project will provide an unparalleled view of proteome organization in a higher eukaryote. The data are being continuously uploaded to the project website, *Drosophila* Protein Interaction Map (DPiM, <https://interfly.med.harvard.edu>), and are integrated into FlyBase. In another large-scale AP-MS project, the Perrimon lab has recently used the TAP approach for mapping protein complexes and probing their dynamics in the receptor tyrosine kinase/Ras/ERK signaling network, using *Drosophila* S2 cells (A. Friedman, N. Perrimon et al., submitted).

6. Mapping post-translational modifications (PTMs) in fly proteins

What is the relationship of the PPI networks generated by the Y2H and AP-MS approaches to the real interactomes in living cells? The PPI networks obtained by high-throughput methods should be interpreted with three caveats: first, such networks represent an aggregate space of all possible PPIs in a cell, while not all of the interactions are realized in a given cell state. Second, these networks represent a static view of the interactome, whereas in reality the interconnections between the proteins are highly dynamic and are under continuous regulation. Finally, PPI networks do not necessarily take into account post-translational protein modifications (PTMs), which can have a major effect on PPIs. Analysis of PTMs thus offers an opportunity to add a dynamic layer to our current view of protein interaction networks.

In the past few years, major advances in this direction have been made in *Drosophila*. Three large-scale studies explored the *Drosophila* phosphoproteome. In the first, Kc167 cells were grown under different conditions (varying nutrient composition of the medium, inclusion of growth stimulants or inhibitors in the medium, and inclusion of phosphatase inhibitors), and phosphorylation sites were determined in the combined peptide sample by mass spectrometry [78]. Sample fractionation and phosphopeptide enrichment are required to maximize the coverage of the phosphoproteome. In the study by Bodenmiller et al., peptides were initially separated by peptide isoelectric focusing, and then three different phosphopeptide isolation methods were used as an enrichment strategy: immobilized metal affinity chromatography (IMAC), titanium dioxide (TiO₂), and phosphoramidate chemistry (PAC). This approach resulted in an identification of 10,118 high confidence phosphorylation sites from 3472 gene models and 4583 distinct phosphoproteins [78]. The data from this project are available via the PhosphoPep database (<http://www.phosphopep.org>). Like the data in

PeptideAtlas, PhosphoPep includes a searchable consensus spectral library that can be useful for designing targeted mass spectrometry experiments. In terms of the functional assignment of the phosphorylated proteins, this study found an enrichment for phosphoproteins involved in regulatory processes, such as kinases, transcription factors, ion channels, and developmental processes. In contrast, phosphoproteins were depleted in the metabolism and metabolic process categories. As an example of the utility of this dataset, the authors analyzed phosphorylation status of the insulin/TOR pathway and found multiple new phosphorylation sites on pathway components [78]. Furthermore, the data from this study was recently used in combination with RNA interference screening to study phosphorylation networks regulating the activity of *Drosophila* JUN N-terminal kinase (JNK) [79].

In another large-scale systematic study, S. Gygi et al. analyzed the phosphoproteome of *Drosophila* embryos [80]. An enrichment strategy used in this study somewhat differed from the one by Bodenmiller et al. and employed strong cation exchange chromatography (SCX) combined with IMAC [80,81]. The authors identified 13,720 unique phosphorylation sites from 2702 proteins with an estimated false discovery rate of 0.63% at the peptide level. Using the Motif-X algorithm, many new consensus phosphorylation motifs were extracted from the dataset. A comparison with the Bodenmiller et al. study revealed relatively low overlap between the two datasets (27% of identified phosphorylation sites and 48% of phosphoproteins). The numbers of multiple phosphorylated proteins were also different (13% in Bodenmiller et al. vs. 68% in Zhai et al.). It is clear that multiple analysis strategies will be required to reveal the full complexity of the *Drosophila* phosphoproteome.

In a third global phosphosite mapping experiment, Hilger et al. detected more than 10,000 phosphorylation sites in *Drosophila* S2 cells and then used this dataset in a quantitative proteomics experiment to study the effect of a phosphatase knockdown on the phosphoproteome [82] (see below). SCX in combination with TiO₂ chromatography was used to enrich for phosphopeptides. There was a good overlap with the other two datasets in terms of identified phosphorylated proteins (65.9% overlap with Zhai et al., and 75.2% with Bodenmiller et al.). At the same time, the Hilger et al. study detected 4691 novel phosphorylation sites, while 5051 (51.8%) were covered by the other two studies, indicating that our knowledge of the *Drosophila* phosphoproteome is not yet exhaustive. This study also performed a bioinformatics analysis of the *Drosophila* phosphoproteome and found that its overall properties are similar to the human phosphoproteome [82]. Moreover, phosphorylation substrates in *Drosophila* can largely be predicted by human kinase motifs, indicating a high degree of conservation of kinases and their signaling pathways. Phosphorylation site data from Hilger et al. are available as part of the Phosida database (<http://www.phosida.com/>).

Two recent studies focused on the proteins phosphorylated on tyrosine residues. Immunoaffinity isolation of the phosphotyrosine subproteome from *Drosophila* S2 cells treated with pervanadate followed by enrichment of phosphopeptides resulted in identification of 562 non-redundant phosphotyrosine sites in 245 proteins [83]. Substrate trapping

was further used to identify potential substrates of protein tyrosine phosphatase dPTP61F [83]. Krishnamoorthy analyzed tyrosine phosphorylation events in *Drosophila* S2 cells after EGFR and insulin-like receptor stimulation [84]. The latter study also identified conserved tyrosine residues in homologous human proteins, suggesting that some of the detected phosphorylation events may be conserved.

In addition to the study of protein phosphorylation, proteomic approaches are being applied to the analysis of other types of PTMs in *Drosophila*. Glycosylation is a PTM that often plays an important regulatory role in controlling protein interactions and activity. Mass spectrometry based approaches have been successfully used to study the structures of the oligosaccharide groups themselves, but the identification of glycosylation sites on proteins is only starting to be explored (reviewed in [85]). A recent study identified 205 glycoproteins carrying N-linked glycans in the *Drosophila* central nervous system, and revealed their 307 N-glycan attachment sites [86]. This study also statistically analyzed amino acid distribution around the N-linked glycosylation sites. Analysis of the mucin-type O-glycoproteome of *Drosophila* S2 cells identified 21 secreted and intracellular glycoproteins [87]. Both of these studies showed that the functional repertoire of the identified glycoproteins is very diverse [85]. Proteomic approaches have also been used in studies focused on glycosylation of individual proteins. For example, a recent report mapped O-mannosylation sites in the *Drosophila* Dystroglycan protein and suggested that *Drosophila* can be a suitable model for studying molecular and genetic mechanisms underlying human dystroglycanopathies [88]. Addition of specific sugars can dramatically change the activity of certain proteins. Thus, glycosylation of the Notch receptor is a critical regulatory mechanism that modulates the affinity and binding preference of the receptor for its ligands (reviewed in [89]). Recent studies of Notch glycosylation have included proteomic experiments to characterize the sites and dynamics of receptor glycosylation [90–92].

Proteomic approaches also make it possible to analyze protein modifications that are difficult to study by other methods. E. Brunner et al. have recently analyzed N-terminal acetylation in *Drosophila* Kc167 cells using combined fractional diagonal chromatography (COFRADIC) followed by mass spectrometry [93]. More than 1200 mature protein N termini were detected, and about 71% of the N-terminal peptides were found to be acetylated, which is close to the reported frequency of N-terminal acetylation in humans. An interesting conclusion from this study is that a proline residue at the first or second position of the mature protein N terminus prevents amino-terminal modification by the acetylation machinery. This property of proline to inhibit N-terminal acetylation, which the authors named the “(X)PX rule”, extends to all species analyzed so far and appears to be a general inhibitory signal. The authors showed that this rule can be used to genetically engineer a protein to study the biological relevance of the presence or absence of an acetyl group, thus providing a generic assay to probe the functional importance of N-terminal acetylation [93]. Proteomics also helps study PTMs of “difficult” proteins, such as histones [94].

In summary, analysis of post-translational protein modifications adds another dimension to our understanding of the

proteome, and reveals an enormous regulatory complexity that underlies protein activities in the cell.

7. Revealing proteome dynamics: quantitative proteomics in *Drosophila*

Classical genetic approaches to studying gene function involve comparisons between the mutant and wild type conditions, using a variety of functional assays such as organism viability or morphology. To understand a global response of the genome to a given perturbation or condition, gene expression analysis via microarrays has been widely used. Yet the effectors of cellular functions are in most cases proteins, and therefore the ability to quantify protein levels on a global scale would provide the most direct means to characterize gene function. In recent years, quantitative proteomics approaches have been applied with success in *Drosophila*, opening a new exciting chapter in fly proteome analysis.

Mass spectrometry is not inherently quantitative, and specialized techniques are required for obtaining information about relative protein abundances. Historically, two-dimensional gel electrophoresis (2-DE) or two-dimensional differential gel electrophoresis (2D-DIGE) followed by mass-spectrometric identification of differentially expressed protein spots has been a method of choice. More recently, protein labeling approaches have been developed that provide an ability to compare protein abundances in two samples by using internal standards in the form of isotopically labeled peptides whose masses are distinguishable in a mass spectrometer. Most approaches involve incorporation of a stable isotope label either *in vivo*, before protein extraction, or *in vitro* via chemical coupling of isotopic tags with predefined masses. Label-free methods are also being developed. The reader is referred to recent reviews on the methodology of quantitative proteomics [1,95]. Below, application of these methods to *Drosophila* proteomics is presented.

The 2-DE and 2D-DIGE methods have been employed in many studies to compare protein levels under different experimental conditions. This method has been particularly successful in characterizing proteins and peptides during the immune response, such as the one induced after infection of *Drosophila* larvae, adults, and cultured cells (reviewed in [96,97]). 2D-DIGE has also been applied to the study of ventral furrow morphogenesis during gastrulation in the embryo [98]. More than 50 proteins with altered abundance levels or isoform changes were identified in genetically ventralized vs. lateralized embryos. A recurring theme in this and other quantitative proteomics studies that compare protein and mRNA levels is a low degree of correlation between the corresponding transcript and protein levels, suggesting that post-translational regulation plays an important role in controlling protein abundance, and that proteomic and transcriptomic approaches are complementary [98–102]. 2-DE was also used to study changes in larval midgut proteins in response to dietary Bowman-Birk inhibitor [103] and to characterize mating-responsive proteins in reproductive tissues in females [99].

A recent introduction of methods that make use of labeling proteins or peptides with stable isotopes provides the most

unbiased and comprehensive way for absolute quantification of proteins in complex samples [95,104]. Advances in mass spectrometry instrumentation and parallel development of software and automation methods for these types of experiments make it possible to quantify complex protein mixtures [105]. *In vitro* (or chemical) labeling approaches, such as isotope-coded affinity tags (ICAT), isotope-coded protein labeling (ICPL), isobaric tags for relative and absolute quantification (iTRAQ), and ^{18}O labeling, introduce an isotopic tag after protein extraction. These techniques have an advantage of being completely independent of the source and preparation of the sample, and are relatively fast. For example, in a study employing ^{18}O and cleavable ICAT (cICAT) labeling of proteins extracted from *Drosophila* cultured cells, the rate of viral protein production was measured during the course of infection with the Flock House Virus (FHV) [106]. In the same study, changes in the host cell proteome were also determined, which resulted in identification of over 200 proteins that were either up- or down-regulated in response to viral infection. A recent report used cICAT labeling and quantitative mass spectrometry to determine the differences in protein expression levels between the fat bodies of normal and starved larvae [107]. Subsequent functional characterization of candidate differentially expressed proteins established the role of the lipid desaturase *Desat1* in starvation-induced autophagy. In another study, four-plex iTRAQ isotope labeling was used in combination with protein complex purification and phosphatase treatment to determine constituents of the protein complex, detect changes in protein composition, and to localize phosphorylation sites and estimate their respective stoichiometry, in a single experiment [108]. This analysis showed that hormone stimulation increases the association of 14-3-3 proteins with the insulin receptor substrate homolog Chico and modulates several phosphorylation sites on the Chico protein. Differential protein labeling with stable isotopes *in vitro* using global internal standard technology (GIST) was applied in a quantitative proteome analysis of a presymptomatic A53T α -synuclein *Drosophila* model of Parkinson disease [109]. A similar approach was later used to study the proteome response to panneuronal expression of human α -synuclein [110]. In another study, a quantitative proteomic analysis of *Drosophila parkin* null mutants and age-matched controls was performed utilizing both GIST and extracted ion chromatogram peak area (XICPA) label-free approaches [111].

An exciting innovation in quantitative proteomics, which is particularly applicable to *Drosophila* as a model organism, is to incorporate stable isotopes into proteins metabolically (i.e. *in vivo*), by growing the organism or cells on special media supplemented with isotopes. The relative abundances of individual proteins obtained from labeled and unlabeled samples can then be directly determined by mass spectrometry. Metabolic labeling of flies was achieved by feeding them uniformly ^{15}N -labeled yeast [112]. Importantly for such type of analysis, a virtually complete labeling was achieved in one generation, without any detrimental effects on the viability of flies or the relative protein abundances [112]. There are distinct advantages to metabolic labeling, compared to the chemical methods: i) labeling is complete, ii) variation in sample handling is eliminated because proteins from

unlabeled and labeled organisms are processed in the same reaction tube, and iii) no derivatization steps are needed after protein extraction. ^{15}N metabolic labeling was used to measure changes in protein expression levels before and after the maternal-to-zygotic transition (MZT) in the *Drosophila* embryos [101]. This study identified with high confidence ~350 proteins that increased in abundance, representing a product of embryonic translation. The group of ~230 down-regulated proteins was dominated by maternal factors involved in translational control. Surprisingly, the mRNA levels of down-regulated proteins remained relatively constant, indicating a translational control mechanism specifically targeting these proteins [101]. An ingenious application of metabolic labeling was used to investigate *Drosophila* seminal fluid proteins transferred at mating ([113,114], reviewed in [115]). In this differential labeling method, female flies were isotopically labeled with ^{15}N and mated with unlabeled males. In subsequent mass spectrometry analysis of proteins from dissected female reproductive tracts female proteins were not identified because the masses of their peptide fragments were increased by incorporation of the heavy nitrogen. These studies identified over 60 proteins not previously known to be involved in reproduction and discovered 38 previously unannotated genes encoding seminal fluid proteins [113,114].

Stable isotope labeling with amino acids in cell culture (SILAC, [116]) has been recently used in *Drosophila* in combination with other approaches to obtain systems-level understanding of gene function. In one study, a combined use of SILAC and RNAi-mediated knockdown of a chromatin remodeling protein ISWI in S2 cells identified ~300 proteins that were significantly up- or down-regulated, about 8% of the detected proteome [102]. Affymetrix-based transcriptomics performed in the same system showed only a limited correlation between mRNA and protein level changes, suggesting an importance of post-transcriptional regulation in determining protein levels in the cell [102]. In another study, SILAC was used in combination with the RNAi-mediated knockdown of phosphatase Ptp61F in S2 cells, and changes in protein phosphorylation were quantified [82]. This study also identified more than 10,000 phosphorylation sites (see above). Interestingly, apart from Ptp61F itself, the proteome was minimally affected by the knockdown, whereas 288 of 6478 high confidence phosphorylation sites changed significantly [82]. This work represents proof of principle that the combination of large-scale phosphoproteomics and a loss of function approach can contribute to elucidating the role of key players in phosphorylation-dependent signaling networks.

8. Summary and outlook

It is clear that in the past decade the young field of *Drosophila* proteomics has blossomed into an active area of research. Proteomics is improving genome annotation and is providing insights into gene function that are not obtainable by other means of analysis. The future of proteomics in flies is bright but there are a few challenges along the way. About a third of the proteome is still undiscovered in shotgun experiments

[33], so additional fractionation methods and possibly instrument improvements are needed before we can “see” every protein in the fly. Resources such as PeptideAtlas, PhosphoPep and Phosida form a foundation for moving from the discovery phase towards directed, hypothesis-driven proteomics [117]. Completion of large-scale PPI studies such as the *Drosophila* Protein Interaction Map will provide a new level of knowledge about protein complexes. Proteome cataloging and annotation efforts will also facilitate comparative studies with other animal groups [34] and with the other sequenced *Drosophila* species [118].

Proteomics, like other high-throughput methods, generates large amounts of data. Presenting and organizing data from proteomics experiments and integrating the results from different types of high-throughput screens remains a challenge. Some of the data, for example from PeptideAtlas, is already being incorporated into FlyBase, but other kinds of databases may provide an additional representation of the data obtained in proteomics studies. While certain databases focus on one aspect of the analysis such as PPIs (e.g. DroID), others synthesize multiple sources of information into one meta-database (e.g. FlyMine, <http://www.flymine.org/>). Proteomics experiments in *Drosophila* are augmented by the resources that emphasize protein localization, such as the FlyTrap (<http://flytrap.med.yale.edu>, [119–121]) and an atlas of gene expression patterns (<http://www.fruitfly.org/insitu>, [122]). *Drosophila* proteomics resources are listed in Table 2, and the reader is further referred to two recent reviews on bioinformatics and proteomics tools for *Drosophila* [123,124], as well as a review by L. Martens in this issue.

One area that should see rapid growth is *Drosophila* quantitative proteomics. The range of experiments that are possible with metabolically labeled flies is now only limited by researcher’s imagination. Quantitative proteomics also offers a means to analyze PPIs, PTMs and other functional relationships in the mutant vs. wild type condition [1]. We should see new functional studies that combine proteomics with the analysis of other types of networks — genetic, phenotypic, transcriptional, and metabolic. Proteomics in *Drosophila* is well under way to help realize the promise of systems biology, which is to define cellular functions for every gene product in the organism.

Table 2 – *Drosophila* proteomics databases.

Purpose of database	Resource	Web page
Central information repository	FlyBase FlyMine	http://flybase.org/ http://www.flymine.org/
Proteome cataloging and annotation	PeptideAtlas PhosphoPep Phosida	http://www.peptideatlas.org/ http://www.phosphopep.org/ http://www.phosida.com/
Protein–protein interactions	DroID BioGRID DPiM	http://www.droidb.org/ http://www.thebiogrid.org/ https://interfly.med.harvard.edu/
Protein expression and localization patterns	FlyTrap In situ atlas of gene expression	http://flytrap.med.yale.edu http://www.fruitfly.org/insitu

Acknowledgements

The author thanks Robert Obar, Jean-François Rual and Marla Tipping for the fruitful discussions and comments on the manuscript. The author extends apologies to the researchers whose work was not cited due to space constraints. A.V. is supported by a grant from the National Science Foundation number 0640700.

REFERENCES

- [1] Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* 2009;10:617–27.
- [2] Beller M, Oliver B. One hundred years of high-throughput *Drosophila* research. *Chromosome Res* 2006;14:349–62.
- [3] Bier E. *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet* 2005;6:9–23.
- [4] Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000;287:2185–95.
- [5] Veraksa A, Del Campo M, McGinnis W. Developmental patterning genes and their conserved functions: from model organisms to humans. *Mol Genet Metab* 2000;69:85–100.
- [6] Ashburner M, Hawley RS, Golic KG. *Drosophila*: a Laboratory Handbook. 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2005
- [7] Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, et al. A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat Biotechnol* 2007;25:576–83.
- [8] Santaren JF. Towards establishing a protein database of *Drosophila*. *Electrophoresis* 1990;11:254–67.
- [9] Taraszka JA, Gao X, Valentine SJ, Sowell RA, Koeniger SL, Miller DF, et al. Proteome profiling for assessing diversity: analysis of individual heads of *Drosophila melanogaster* using LC-ion mobility-MS. *J Proteome Res* 2005;4:1238–47.
- [10] Taraszka JA, Kurulugama R, Sowell RA, Valentine SJ, Koeniger SL, Arnold RJ, et al. Mapping the proteome of *Drosophila melanogaster*: analysis of embryos and adult heads by LC-IMS-MS methods. *J Proteome Res* 2005;4:1223–37.
- [11] Takemori N, Yamamoto MT. Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* 2009;9:2484–93.
- [12] Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat Genet* 2006;38:1440–5.
- [13] Karr TL. Fruit flies and the sperm proteome. *Hum Mol Genet* 2007;16 Spec no. 2: R124–33.
- [14] Alonso J, Santaren JF. Proteomic analysis of the wing imaginal discs of *Drosophila melanogaster*. *Proteomics* 2005;5:474–89.
- [15] Guedes Sde M, Vitorino R, Tomer K, Domingues MR, Correia AJ, Amado F, et al. *Drosophila melanogaster* larval hemolymph protein mapping. *Biochem Biophys Res Commun* 2003;312: 545–54.
- [16] de Moraes Guedes S, Vitorino R, Domingues R, Tomer K, Correia AJ, Amado F, et al. Proteomics of immune-challenged *Drosophila melanogaster* larvae hemolymph. *Biochem Biophys Res Commun* 2005;328:106–15.
- [17] Vierstraete E, Cerstiaens A, Baggerman G, Van den Bergh G, De Loof A, Schoofs L. Proteomics in *Drosophila melanogaster*: first 2D database of larval hemolymph proteins. *Biochem Biophys Res Commun* 2003;304:831–8.
- [18] Vierstraete E, Verleyen P, Baggerman G, D’Hertog W, Van den Bergh G, Arckens L, et al. A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. *Proc Natl Acad Sci U S A* 2004;101:470–5.
- [19] Karlsson C, Korayem AM, Scherfer C, Loseva O, Dushay MS, Theopold U. Proteomic analysis of the *Drosophila* larval hemolymph clot. *J Biol Chem* 2004;279:52033–41.
- [20] Levy F, Bulet P, Ehret-Sabatier L. Proteomic analysis of the systemic immune response of *Drosophila*. *Mol Cell Proteomics* 2004;3:156–66.
- [21] Verleyen P, Baggerman G, D’Hertog W, Vierstraete E, Husson SJ, Schoofs L. Identification of new immune induced molecules in the haemolymph of *Drosophila melanogaster* by 2D-nanoLC MS/MS. *J Insect Physiol* 2006;52:379–88.
- [22] Baggerman G, Boonen K, Verleyen P, De Loof A, Schoofs L. Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J Mass Spectrom* 2005;40:250–60.
- [23] Baggerman G, Cerstiaens A, De Loof A, Schoofs L. Peptidomics of the larval *Drosophila melanogaster* central nervous system. *J Biol Chem* 2002;277:40368–74.
- [24] Schoofs L, Baggerman G. Peptidomics in *Drosophila melanogaster*. *Brief Funct Genomic Proteomic* 2003;2:114–20.
- [25] Predel R, Wegener C, Russell WK, Tichy SE, Russell DH, Nachman RJ. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: a mass spectrometric survey of peptides from individual flies. *J Comp Neurol* 2004;474:379–92.
- [26] Wegener C, Reintl T, Jansch L, Predel R. Direct mass spectrometric peptide profiling and fragmentation of larval peptide hormone release sites in *Drosophila melanogaster* reveals tagma-specific peptide expression and differential processing. *J Neurochem* 2006;96:1362–74.
- [27] Neupert S, Johard HA, Nassel DR, Predel R. Single-cell peptidomics of *Drosophila melanogaster* neurons identified by Gal4-driven fluorescence. *Anal Chem* 2007;79:3690–4.
- [28] Alonso J, Rodriguez JM, Baena-Lopez LA, Santaren JF. Characterization of the *Drosophila melanogaster* mitochondrial proteome. *J Proteome Res* 2005;4:1636–45.
- [29] Alonso J, Santaren JF. Characterization of the *Drosophila melanogaster* ribosomal proteome. *J Proteome Res* 2006;5: 2025–32.
- [30] Cermelli S, Guo Y, Gross SP, Welte MA. The lipid-droplet proteome reveals that droplets are a protein-storage depot. *Curr Biol* 2006;16:1783–95.
- [31] Beller M, Riedel D, Jansch L, Dieterich G, Wehland J, Jackle H, et al. Characterization of the *Drosophila* lipid droplet subproteome. *Mol Cell Proteomics* 2006;5:1082–94.
- [32] Ahrens CH, Brunner E, Hafen E, Aebersold R, Basler K. A proteome catalog of *Drosophila melanogaster*: an essential resource for targeted quantitative proteomics. *Fly (Austin)* 2007;1:182–6.
- [33] Loevenich SN, Brunner E, King NL, Deutsch EW, Stein SE, Aebersold R, et al. The *Drosophila melanogaster* PeptideAtlas facilitates the use of peptide data for improved fly proteomics and genome annotation. *BMC Bioinformatics* 2009;10:59.
- [34] Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmstrom J, et al. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol* 2009;7:e48.
- [35] Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmens I, Hirozane-Kishikawa T, et al. Empirically controlled mapping of the *Caenorhabditis elegans* protein–protein interactome network. *Nat Methods* 2009;6:47–54.
- [36] Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 2009;10:2763–88.
- [37] Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 2007;8:645–54.

- [38] Kocher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 2007;4:807–15.
- [39] Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interaction map of *Drosophila melanogaster*. *Science* 2003;302:1727–36.
- [40] Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, et al. An empirical framework for binary interactome mapping. *Nat Methods* 2009;6:83–90.
- [41] Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, Trehin A, et al. Protein interaction mapping: a *Drosophila* case study. *Genome Res* 2005;15:376–84.
- [42] Stanyon CA, Liu G, Mangiola BA, Patel N, Giot L, Kuang B, et al. A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol* 2004;5:R96.
- [43] Schwartz AS, Yu J, Gardenour KR, Finley Jr RL, Ideker T. Cost-effective strategies for completing the interactome. *Nat Methods* 2009;6:55–61.
- [44] Yu J, Pacifico S, Liu G, Finley Jr RL. DroID: the *Drosophila* Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics* 2008;9:461.
- [45] Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, et al. High-quality binary protein interaction map of the yeast interactome network. *Science* 2008;322:104–10.
- [46] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* 2006;440:631–6.
- [47] Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 2003;60:523–33.
- [48] Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein Expr Purif* 2005;41:98–105.
- [49] Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 2001;412:655–60.
- [50] Schaefer M, Shevchenko A, Knoblich JA. A protein complex containing Inscuteable and the Galpha-binding protein Pins orients asymmetric cell divisions in *Drosophila*. *Curr Biol* 2000;10:353–62.
- [51] Furuyama T, Henikoff S. Biotin-tag affinity purification of a centromeric nucleosome assembly complex. *Cell Cycle* 2006;5:1269–74.
- [52] Kroiss M, Brunger KM, Wiesner J, Grimm M, Sickmann A, Fischer U. Native purification of protein and RNA–protein complexes using a novel affinity procedure. *Fly (Austin)* 2009;3:221–8.
- [53] Schmidt TG, Skerra A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* 2007;2:1528–35.
- [54] Collins MO, Choudhary JS. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr Opin Biotechnol* 2008;19:324–30.
- [55] Glatter T, Wepf A, Aebersold R, Gstaiger M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol Syst Biol* 2009;5:237.
- [56] Yang P, Sampson HM, Krause HM. A modified tandem affinity purification strategy identifies cofactors of the *Drosophila* nuclear receptor dHNF4. *Proteomics* 2006;6:927–35.
- [57] Xu X, Song Y, Li Y, Chang J, Zhang H, An L. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. *Protein Expr Purif* 2010;72:149–56.
- [58] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 1999;17:1030–2.
- [59] Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;415:141–7.
- [60] Westermarck J, Weiss C, Saffrich R, Kast J, Musti AM, Wessely M, et al. The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J* 2002;21:451–60.
- [61] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006;440:637–43.
- [62] Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* 2004;6:97–105.
- [63] Forler D, Kocher T, Rode M, Gentzel M, Izaurralde E, Wilm M. An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat Biotechnol* 2003;21:89–92.
- [64] Veraksa A, Bauer A, Artavanis-Tsakonas S. Analyzing protein complexes in *Drosophila* with tandem affinity purification-mass spectrometry. *Dev Dyn* 2005;232:827–34.
- [65] Moberg KH, Mukherjee A, Veraksa A, Artavanis-Tsakonas S, Hariharan IK. The *Drosophila* F box protein archipelago regulates dMyc protein levels in vivo. *Curr Biol* 2004;14:965–74.
- [66] Mukherjee A, Veraksa A, Bauer A, Rosse C, Camonis J, Artavanis-Tsakonas S. Regulation of Notch signalling by non-viral beta-arrestin. *Nat Cell Biol* 2005;7:1191–201.
- [67] Hurt JA, Obar RA, Zhai B, Farny NG, Gygi SP, Silver PA. A conserved CCCH-type zinc finger protein regulates mRNA nuclear adenylation and export. *J Cell Biol* 2009;185:265–77.
- [68] Zhou R, Czech B, Brennecke J, Sachidanandam R, Wohlschlegel JA, Perrimon N, et al. Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA* 2009;15:1886–95.
- [69] Wainman A, Creque J, Williams B, Williams EV, Bonaccorsi S, Gatti M, et al. Roles of the *Drosophila* NudE protein in kinetochore function and centrosome migration. *J Cell Sci* 2009;122:1747–58.
- [70] Burgio G, La Rocca G, Sala A, Arancio W, Di Gesu D, Collesano M, et al. Genetic identification of a network of factors that functionally interact with the nucleosome remodeling ATPase ISWI. *PLoS Genet* 2008;4:e1000089.
- [71] Johnson RI, Seppa MJ, Cagan RL. The *Drosophila* CD2AP/CIN85 orthologue Cindr regulates junctions and cytoskeleton dynamics during tissue patterning. *J Cell Biol* 2008;180:1191–204.
- [72] Williams B, Leung G, Maiato H, Wong A, Li Z, Williams EV, et al. Mitch a rapidly evolving component of the Ndc80 kinetochore complex required for correct chromosome segregation in *Drosophila*. *J Cell Sci* 2007;120:3522–33.
- [73] Lake RJ, Grimm LM, Veraksa A, Banos A, Artavanis-Tsakonas S. In vivo analysis of the Notch receptor S1 cleavage. *PLoS One* 2009;4:e6728.
- [74] Omori Y, Zhao C, Saras A, Mukhopadhyay S, Kim W, Furukawa T, et al. Elipsa is an early determinant of ciliogenesis that links the IFT particle to membrane-associated small GTPase Rab8. *Nat Cell Biol* 2008;10:437–44.
- [75] Burckstummer T, Bennett KL, Preradovic A, Schutze G, Hantschel O, Superti-Furga G, et al. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods* 2006;3:1013–9.
- [76] Kyriakakis P, Tipping M, Abed L, Veraksa A. Tandem affinity purification in *Drosophila*: the advantages of the GS-TAP system. *Fly (Austin)* 2008;2:229–35.
- [77] Keefe AD, Wilson DS, Seelig B, Szostak JW. One-step purification of recombinant proteins using a

- nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein Expr Purif* 2001;23:440–6.
- [78] Bodenmiller B, Malmstrom J, Gerrits B, Campbell D, Lam H, Schmidt A, et al. PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol Syst Biol* 2007;3:139.
- [79] Bakal C, Linding R, Llense F, Heffern E, Martin-Blanco E, Pawson T, et al. Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 2008;322:453–6.
- [80] Zhai B, Villen J, Beausoleil SA, Mintseris J, Gygi SP. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res* 2008;7:1675–82.
- [81] Villen J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 2008;3:1630–8.
- [82] Hilger M, Bonaldi T, Gnad F, Mann M. Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol Cell Proteomics* 2009;8:1908–20.
- [83] Chang YC, Lin SY, Liang SY, Pan KT, Chou CC, Chen CH, et al. Tyrosine phosphoproteomics and identification of substrates of protein tyrosine phosphatase dPTP61F in *Drosophila* S2 cells by mass spectrometry-based substrate trapping strategy. *J Proteome Res* 2008;7:1055–66.
- [84] Krishnamoorthy S. Receptor tyrosine kinase (RTK) mediated tyrosine phosphor-proteome from *Drosophila* S2 (ErbB1) cells reveals novel signaling networks. *PLoS One* 2008;3:e2877.
- [85] ten Hagen KG, Zhang L, Tian E, Zhang Y. Glycobiology on the fly: developmental and mechanistic insights from *Drosophila*. *Glycobiology* 2009;19:102–11.
- [86] Koles K, Lim JM, Aoki K, Porterfield M, Tiemeyer M, Wells L, et al. Identification of N-glycosylated proteins from the central nervous system of *Drosophila melanogaster*. *Glycobiology* 2007;17:1388–403.
- [87] Schwientek T, Mandel U, Roth U, Muller S, Hanisch FG. A serial lectin approach to the mucin-type O-glycoproteome of *Drosophila melanogaster* S2 cells. *Proteomics* 2007;7:3264–77.
- [88] Nakamura N, Stalnaker SH, Lyalin D, Lavrova O, Wells L, Panin VM. *Drosophila* Dystroglycan is a target of O-mannosyltransferase activity of two protein O-mannosyltransferases, rotated abdomen and twisted. *Glycobiology* 2010;20:381–94.
- [89] Stanley P. Regulation of Notch signaling by glycosylation. *Curr Opin Struct Biol* 2007;17:530–5.
- [90] Xu A, Haines N, Dlugosz M, Rana NA, Takeuchi H, Haltiwanger RS, et al. In vitro reconstitution of the modulation of *Drosophila* Notch-ligand binding by Fringe. *J Biol Chem* 2007;282:35153–62.
- [91] Acar M, Jafar-Nejad H, Takeuchi H, Rajan A, Ibrani D, Rana NA, et al. Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* 2008;132:247–58.
- [92] Matsuura A, Ito M, Sakaidani Y, Kondo T, Murakami K, Furukawa K, et al. O-linked N-acetylglucosamine is present on the extracellular domain of notch receptors. *J Biol Chem* 2008;283:35486–95.
- [93] Goetze S, Qeli E, Mosimann C, Staes A, Gerrits B, Roschitzki B, et al. Identification and functional characterization of N-terminally acetylated proteins in *Drosophila melanogaster*. *PLoS Biol* 2009;7:e1000236.
- [94] Bonaldi T, Imhof A, Regula JT. A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics* 2004;4:1382–96.
- [95] Bachi A, Bonaldi T. Quantitative proteomics as a new piece of the systems biology puzzle. *J Proteomics* 2008;71:357–67.
- [96] Engstrom Y, Loseva O, Theopold U. Proteomics of the *Drosophila* immune response. *Trends Biotechnol* 2004;22:600–5.
- [97] Vierstraete E, Verleyen P, De Loof A, Schoofs L. Differential proteomics for studying *Drosophila* immunity. *Ann N Y Acad Sci* 2005;1040:504–7.
- [98] Gong L, Puri M, Unlu M, Young M, Robertson K, Viswanathan S, et al. *Drosophila* ventral furrow morphogenesis: a proteomic analysis. *Development* 2004;131:643–56.
- [99] Mack PD, Kapelnikov A, Heifetz Y, Bender M. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 2006;103:10358–63.
- [100] Xun Z, Sowell RA, Kaufman TC, Clemmer DE. Protein expression in a *Drosophila* model of Parkinson's disease. *J Proteome Res* 2007;6:348–57.
- [101] Gouw JW, Pinkse MW, Vos HR, Moshkin Y, Verrijzer CP, Heck AJ, et al. In vivo stable isotope labeling of fruit flies reveals post-transcriptional regulation in the maternal-to-zygotic transition. *Mol Cell Proteomics* 2009;8:1566–78.
- [102] Bonaldi T, Straub T, Cox J, Kumar C, Becker PB, Mann M. Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol Cell* 2008;31:762–72.
- [103] Li HM, Margam V, Muir WM, Murdock LL, Pittendrigh BR. Changes in *Drosophila melanogaster* midgut proteins in response to dietary Bowman-Birk inhibitor. *Insect Mol Biol* 2007;16:539–49.
- [104] Gouw JW, Krijgsveld J, Heck AJ. Quantitative proteomics by metabolic labeling of model organisms. *Mol Cell Proteomics* 2010;9:11–24.
- [105] Bakalarski CE, Elias JE, Villen J, Haas W, Gerber SA, Everley PA, et al. The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J Proteome Res* 2008;7:4756–65.
- [106] Go EP, Wikoff WR, Shen Z, O'Maille G, Morita H, Conrads TP, et al. Mass spectrometry reveals specific and global molecular transformations during viral infection. *J Proteome Res* 2006;5:2405–16.
- [107] Kohler K, Brunner E, Guan XL, Boucke K, Greber UF, Mohanty S, et al. A combined proteomic and genetic analysis identifies a role for the lipid desaturase *Desat1* in starvation-induced autophagy in *Drosophila*. *Autophagy* 2009;5:980–90.
- [108] Pflieger D, Junger MA, Muller M, Rinner O, Lee H, Gehrig PM, et al. Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Mol Cell Proteomics* 2008;7:326–46.
- [109] Xun Z, Sowell RA, Kaufman TC, Clemmer DE. Quantitative proteomics of a presymptomatic A53T alpha-synuclein *Drosophila* model of Parkinson disease. *Mol Cell Proteomics* 2008;7:1191–203.
- [110] Xun Z, Kaufman TC, Clemmer DE. Proteome response to the panneural expression of human wild-type alpha-synuclein: a *Drosophila* model of Parkinson's disease. *J Proteome Res* 2008;7:3911–21.
- [111] Xun Z, Kaufman TC, Clemmer DE. Stable isotope labeling and label-free proteomics of *Drosophila* parkin null mutants. *J Proteome Res* 2009;8:4500–10.
- [112] Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, et al. Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* 2003;21:927–31.
- [113] Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res* 2009;19:886–96.
- [114] Findlay GD, Yi X, MacCoss MJ, Swanson WJ. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* 2008;6:e178.
- [115] Findlay GD, Swanson WJ. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *Bioessays* 2010;32:26–36.

- [116] Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006;7:952–8.
- [117] Schmidt A, Claassen M, Aebersold R. Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* 2009;13:510–7.
- [118] Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 2007;450:203–18.
- [119] Kelso RJ, Buszczak M, Quinones AT, Castiblanco C, Mazzalupo S, Cooley L. Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster*. *Nucleic Acids Res* 2004;32:D418–20.
- [120] Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, et al. The Carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* 2007;175:1505–31.
- [121] Quinones-Coello AT, Petrella LN, Ayers K, Melillo A, Mazzalupo S, Hudson AM, et al. Exploring strategies for protein trapping in *Drosophila*. *Genetics* 2007;175:1089–104.
- [122] Tomancak P, Berman BP, Beaton A, Weizmann R, Kwan E, Hartenstein V, et al. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* 2007;8:R145.
- [123] Costello JC, Cash AC, Dalkilic MM, Andrews JR. Data pushing: a fly-centric guide to bioinformatics tools. *Fly (Austin)* 2008;2:1–18.
- [124] Carmena A. Approaching *Drosophila* development through proteomic tools and databases: at the hub of the post-genomic era. *Mech Dev* 2009;126:761–70.