Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles

Anne-Claude Gingras[1], Kento T Abe[1] and Brian Raught[2]

The use of proximity-dependent biotinylation approaches combined with mass spectrometry (e.g. BioID and APEX) has revolutionized the study of protein–protein interactions and organellar proteomics. These powerful techniques are based on the fusion of an enzyme (e.g. a biotin ligase or peroxidase) to a 'bait' protein of interest, which is then expressed in a relevant biological setting. Addition of enzyme substrate enables covalent biotin labeling of proteins in the vicinity of the bait *in vivo*. These approaches thus allow for the capture and identification of 'neighborhood' proteins in the context of a living cell, and provide data that are complementary to more established techniques such as fractionation or affinity purification. As compared to standard affinity-based purification approaches, proximity-dependent biotinylation (PDB) can help to: first, identify interactions with and amongst membrane proteins, and other polypeptide classes that are less amenable to study by standard pulldown techniques; second, enrich for transient and/or low affinity interactions that are not readily captured using affinity purification approaches; third, avoid post-lysis artefacts associated with standard biochemical purification experiments and; fourth, provide deep insight into the organization of membrane-less organelles and other subcellular structures that cannot be easily isolated or purified. Given the increasing use of these techniques to answer a variety of different types of biological questions, it is important to understand how best to design PDB–MS experiments, what type of data they generate, and how to analyze and interpret the results.

Addresses
[1] Lunenfeld-Tanenbaum Research Institute, Sinai Health System, and Department of Molecular Genetics, University of Toronto, 600 University Ave, Rm 992, Toronto, ON, M5G1X5, Canada
[2] Princess Margaret Cancer Centre, University Health Network, and Department of Medical Biophysics, University of Toronto, 101 College Street, 9-701A, Toronto, ON, M5G 1L7, Canada

Corresponding author: Gingras, Anne-Claude (gingras@lunenfeld.ca)

Current Opinion in Chemical Biology 2019, **48**:44–54

This review comes from a themed issue on **Omics**

Edited by **Ileana M Cristea** and **Kathryn S Lilley**

For a complete overview see the Issue and the Editorial

Available online 17th November 2018

https://doi.org/10.1016/j.cbpa.2018.10.017

Introduction

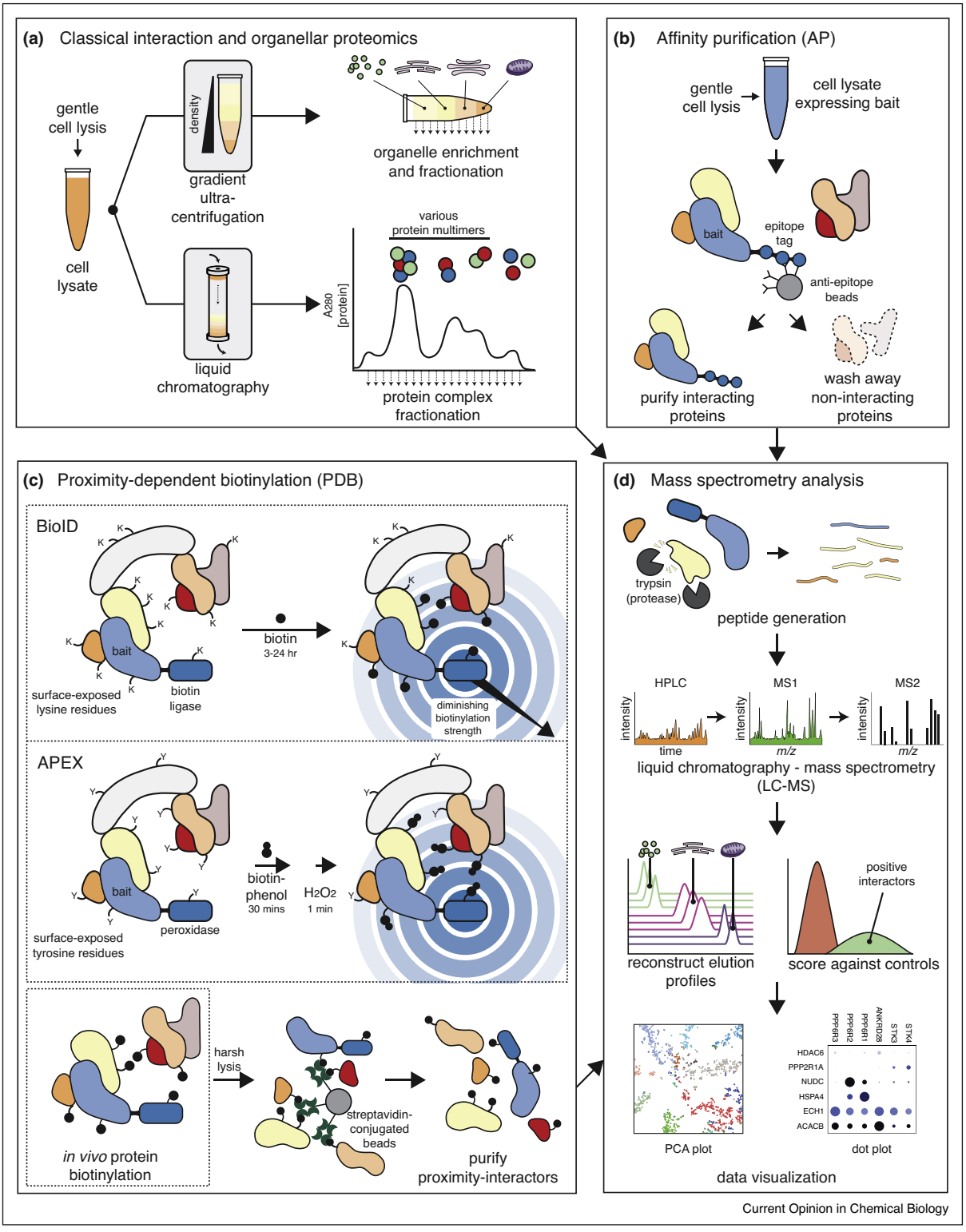
Biological functions are generally performed by proteins as components of complexes, organelles or other assemblies. While the human genome project has provided a 'parts list' of the >20 000 polypeptides putatively encoded in the genome, how these proteins are organized to perform the intricate functions required for life remains incompletely understood. To fill these gaps, a wide variety of biochemical techniques have been developed to characterize protein localization and identify interacting protein partners, including organelle purification and immunoprecipitation coupled with mass spectrometry.

Biochemical fractionation techniques are commonly employed for the characterization of large protein complexes and defining organelle composition (Figure 1a). In organellar proteomics, fractionation approaches such as density gradients are often coupled with mass spectrometric characterization of the fractions (reviewed in Ref. [1]). The incorporation of quantitative proteomics has enabled robust definition of the composition of membrane-bound subcellular structures in diverse biological settings (e.g. [2,3]). Fractionation approaches can similarly be employed to purify biochemically stable protein complexes based on co-fractionation using for example size or charge-based separation techniques [4–6]. Both organellar fractionation and complexome profiling can also be coupled with perturbations such as a drug/hormone treatment or protein knockdown/knockout [6,7].

Complementing these global profiling techniques, common approaches for the identification of protein–protein interactions typically rely on the in-frame fusion of an epitope tag that can be used for purification of a 'bait' polypeptide of interest, or bait-specific affinity reagents such as antibodies (Figure 1b). In these 'affinity purification coupled with mass spectrometry' (AP–MS) techniques, endogenous proteins that interact with the bait can be captured, identified and quantified. AP–MS is widely used to identify interaction partners for a bait polypeptide of interest, and the same approach conducted in an iterative manner can be used to characterize protein complexes, as we discussed previously [8].

While these approaches are extremely powerful, they also share some important limitations. For example, the organelles or complexes under study must first be efficiently extracted during cell or tissue lysis, and then maintain

Figure 1



Organellar and interaction proteomics approaches. **(a)** Schematic of classical organellar (top) and protein complex (bottom) fractionation approaches. A cell is first lysed, and the prepared lysate separated based on biophysical properties of its constituents. For comprehensive characterization, the entire elution profile is analyzed by quantitative MS (see d). **(b)** Standard affinity purification (AP). A bait is often fused to an epitope tag (here, the common 3x-FLAG tag is shown) and expressed in a relevant setting. After cell lysis under mild conditions, the epitope tag is captured on an affinity support, resulting in the purification of the bait and its stable interaction partners. **(c)** In proximity-dependent biotinylation

their integrity throughout subsequent purification, fractionation and washing steps. This poses significant challenges for the characterization of membrane-less organelles or protein complexes localized to poorly soluble subcellular compartments such as membranes, chromatin, nuclear lamina or the cytoskeleton. Many protein–protein interactions can be disrupted under the detergent/buffer conditions required to solubilize proteins from these and other compartments. Similarly, maintaining the integrity of membrane-less organelles can be difficult, and how accurately the resulting purified products reflect the original structures *in vivo* can be difficult to assess. A second issue with standard biochemical approaches is that strong interactors are much easier to capture using these methods than weaker or more transient binding partners. Weak protein–protein interactions play critically important roles in biology [9], but they are most likely significantly underrepresented in current protein interaction databases. A third limitation to these approaches is post-lysis artefacts. Proteins not normally localized at the same intracellular location can be mixed together upon lysis. Interacting partners can thus be extensively 'shuffled' during lysis and affinity purification steps, resulting in the identification of interactions that may not normally take place *in vivo* [10].

Here we describe how recently developed proximity-dependent biotinylation approaches can overcome some of these limitations and be used to identify protein interactors and components of organelles and other subcellular structures. We also include some suggestions for designing more informative experiments, and for the analysis of data resulting from proximity-dependent biotinylation studies.

Proximity-dependent biotinylation coupled with mass spectrometry (PDB–MS)

Limitations associated with classic fractionation or AP–MS approaches have prompted the development of alternative methods, including proximity-dependent biotinylation coupled with mass spectrometry (PDB–MS). These approaches generally involve the fusion of a bait protein with an enzyme that can catalyze the activation of biotin or a phenolic biotin derivative. Expression of this fusion protein in a relevant biological setting, and supplying the relevant biotin substrate, results in the biotinylation of nearby proteins in living cells. Since the biotin 'tag' is covalently attached, the cells/tissue expressing the bait protein can be lysed under harsh buffer conditions to disrupt and efficiently solubilize membranes, organelles and protein complexes. Biotin-tagged proteins are then captured, most often using streptavidin linked to a solid-phase support, and identified by mass spectrometry (Figure 1c).

To date, two main classes of enzymes have been employed for proximity-dependent biotinylation: biotin ligases (BirA* [11••], BioID2 [12], BASU [13], TurboID [14•], miniTurbo [14•]) and peroxidases (APEX [15••], APEX2 [16], HRP [17••,18,19]). Biotin ligases activate biotin to the reactive biotinoyl-AMP intermediate, but have been modified from the wild type enzymes such that they are unable to catalyze the direct transfer of activated biotin to a specific amino acid sequence in a substrate protein. As such, these abortive enzymes simply release a 'cloud' of activated biotin, which can react with epsilon amine groups on lysine residues in nearby polypeptides (Figure 1d). Peroxidases catalyze tyrosine attack by a phenolic compound coupled to biotin (e.g. biotin-phenol), when activated by peroxide. A major difference between the initial BioID biotin ligase (BirA*) and peroxidases (APEX, APEX2) was the length of time required for the generation of sufficient biotinylation signal to be useful for mass spectrometry-based identification approaches. BirA* generally requires several hours of labeling (most published protocols use 12–24 h, though we have employed as low as 3 h [20•]). By contrast, APEX2 labeling requires only minutes of labeling, which has enabled time-sensitive studies, such as the characterization of G-protein-coupled receptor signaling [21•,22•]. The recent development of molecularly evolved versions of biotin ligases that can generate strong signal-to-noise within minutes in cultured cells, such as TurboID and miniTurbo [14•], may soon allow biotin ligases to also be used for the study of more dynamic processes.

PDB–MS—by the numbers

The 'cloud' of activated biotin surrounding the bait protein resembles a contour map, with strongest labeling closer to the enzyme, and weaker labeling as the distance from the bait protein increases. However, the effective radius of labeling likely varies widely depending on the makeup and dynamics of the subcellular structure(s) under study. For instance, a membrane protein of the nuclear envelope could display extensive lateral movement across one face of the lipid bilayer, resulting in very different labeling dynamics than protein components of more stable structures. Using PDB–

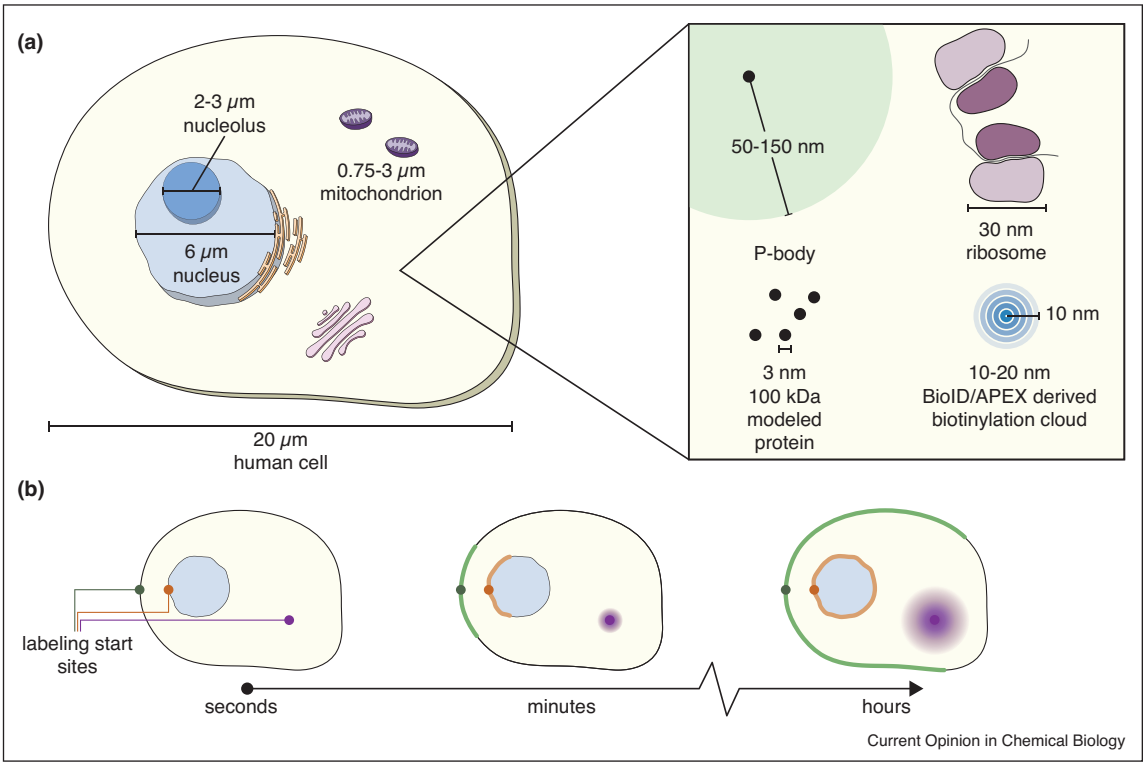
(Figure 1 Legend Continued) (PDB), an enzyme, such as a biotin ligase (e.g. for BioID) or a peroxidase (e.g. in APEX) is fused to the bait of interest and expressed in a relevant setting. Addition of the enzyme substrate (and activator for peroxidase) enable the covalent tagging of reactive residues, namely lysines for BioID and tyrosines for APEX. After cell lysis in harsh (often denaturing) conditions, the biotinylated proteins are recovered by affinity reagents. **(d)** Common to all purification schemes, the fractionated or purified proteins are proteolyzed (e.g. by trypsin) and the peptides usually analyzed by quantitative LC–MS/MS. Depending on the experimental set-up, different data analysis schemes can be used, notably reconstruction of the elution profiles by machine learning and/or correlation analysis (e.g. for panel a), or scoring against sets of negative controls (generally for panels b and c).

MS approaches as 'molecular rulers' can thus be problematic. Despite these caveats, an estimation of the BirA* labeling radius based on the well-characterized organization of the nuclear pore complex is ~10 nm [23] (Figure 2a), and immuno-gold labeling suggested a ~20 nm labeling radius for peroxidases. To put these numbers in context, a HeLa cell is generally 20–40 μm in diameter (see Refs. [24] and http://bionumbers.hms.harvard.edu/search.aspx for size estimations), mitochondria commonly range from 0.75–3 μm in diameter, and membrane-less organelles can range from 100 to 300 nm (P-bodies; [25]) to 2–3 μm (nucleolus). Eukaryotic ribosomes are 30 nm in diameter, and the average globular protein of 30–200 kDa spans 5–10 nm [26]. Thus, the labeling radius of BioID/APEX favors proteins that are in close proximity to the bait, including both direct binding partners and nearby components of protein complexes in which it resides. Importantly, proximity-dependent labeling can also provide valuable information on the immediate intracellular neighborhood occupied by a bait protein; for example if a protein is localized throughout the entire mitochondrial matrix, essentially all accessible matrix proteins have a chance to be labeled. It is however important to keep in mind that subcellular structures are not static; as labeling duration increases, dynamic structures (shown for membranes in Figure 2b) can become progressively labeled over time, dependent on the movement and half-lives of the bait protein and its biotinylated partners.

Characterizing organellar composition using proximity-dependent biotinylation

The first manuscript describing the use of BioID used lamin A as a bait to characterize the composition of the nuclear lamina, a relatively insoluble structure that was historically difficult to study using standard biochemical methods [11••]. Similarly, the initial APEX study used mitochondrial targeting sequences to direct the peroxidase to the matrix and inner mitochondrial space, then used quantitative proteomics approaches to define the composition of these structures, without the need to purify mitochondria [15••]. On the basis of these groundbreaking studies demonstrating that the approaches were capable of overcoming common limitations of standard biochemical purification strategies, other groups have since explored the use of PDB–MS to identify new components of organelles and other subcellular structures (e.g. [17••,20•,27••,28–30,31•,32,33]). Recent examples include the use of APEX2 for profiling the composition of the lipid droplet proteome [34], and for defining

Figure 2



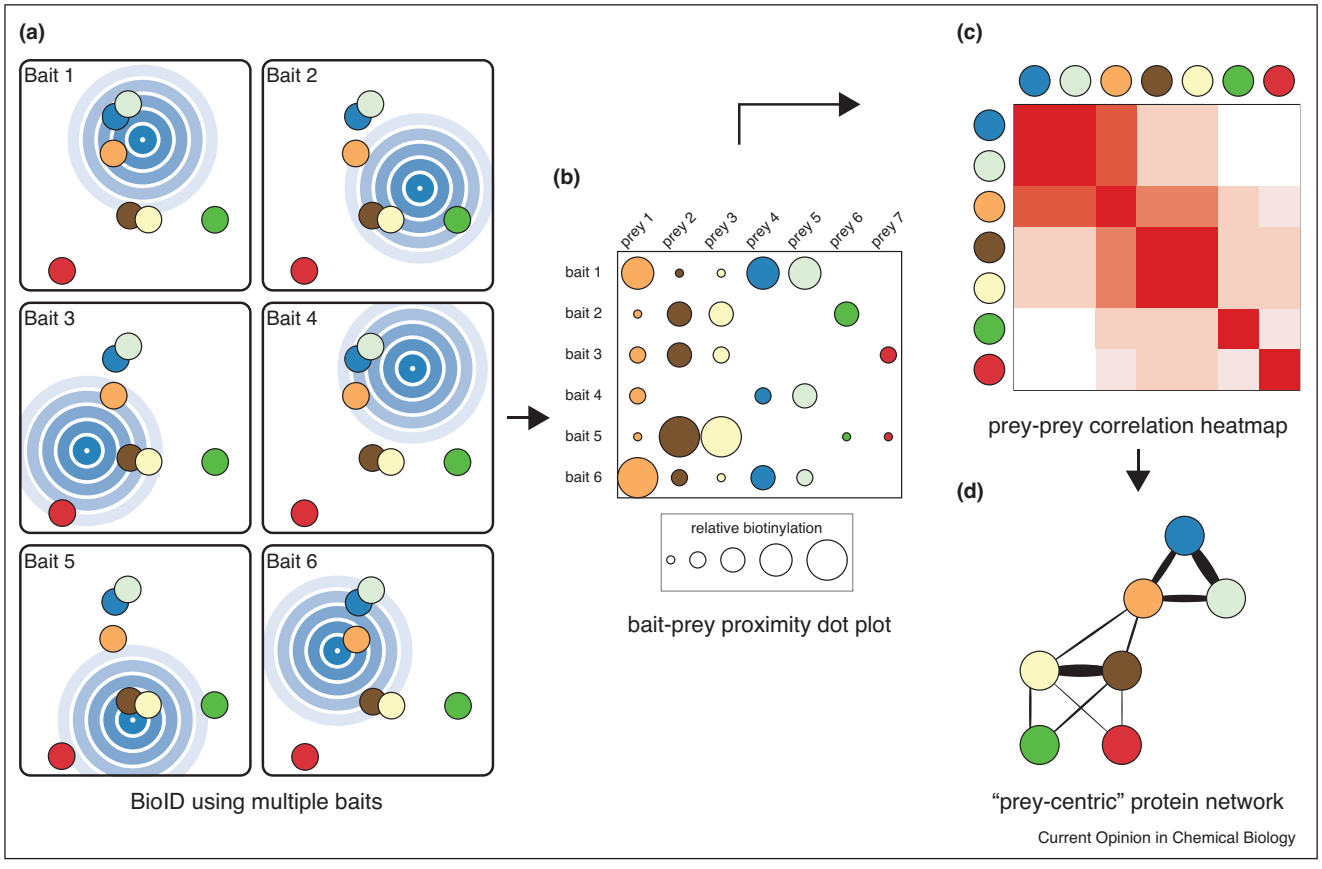
Size and time considerations in PDB–MS experiments. **(a)** Relative diameters of different subcellular structures [24] in relation to the estimated labeling radii of BioID and APEX. **(b)** Consideration of bait and biotinylated prey protein dynamics as a function of time in PDB–MS experiments.

context-dependent association of proteins with the stress granule marker G3BP1 [31•].

While many of the studies listed above used only a few baits to infer the composition of organelles, our own work has demonstrated the advantages of more systematic mapping to define both the composition and organization of membrane-less organelles. In the first large-scale membrane-less organellar study using BioID (combined with a complementary FLAG AP–MS analysis), 58 different baits mapping to the centrosome and primary cilia were used to better define the composition and organization of these important structures [27••]. A network of >7,000 high-confidence proximity interactions amongst 1405 proteins was analyzed using bait-prey hierarchical clustering to reveal a 'core' network of ~540 proteins detected by multiple baits in the dataset. Additional subgroups of baits/preys could also be distinguished using this approach, revealing the organization of previously poorly understood centrosome/cilia substructures. Extensive validation of the localization of newly discovered centrosome-associated preys, and the functional consequences of their depletion on centrosome function, served as confirmation that this methodology can yield key information regarding the composition and organization of membrane-less organelles.

More recently, we used BioID to define the composition and organization of two closely apposed cytosolic RNA-associated structures, P-bodies and stress granules [20•]. Using 139 baits, we generated a bait-prey interactome connecting 1792 proteins via 7424 unique proximity interactions. In this case, we postulated that prey proteins in close proximity to one another should be co-labeled by the same set of baits, allowing for a 'prey-centric' view of the network (Figure 3). As opposed to a standard 'bait-centric' view, this strategy is less reliant on information from individual bait proteins, and can reveal significant additional substructural details of interest: that is revealing preys that are 'closer' to one another, including components of protein complexes and direct interactors. When combined with

Figure 3



Use of PDB–MS to infer proximity relationships between prey proteins. **(a)** Example labeling of seven prey proteins exhibiting to one another a certain distance relationship by six different BioID baits. **(b)** Representation of the quantitative recovery of each prey across all bait purifications. Circle size relates to labeling quantity by each bait. **(c)** Correlation analysis of the quantitative behavior of the preys across the dataset. Color intensity reflects the correlation score. **(d)** Network visualization of prey-prey correlations. Thickness of the edges maps to the correlation coefficient, and only those pairs with a correlation cutoff above an arbitrary threshold are displayed.

computational approaches such as a simple Pearson correlation analysis or categorization using non-negative matrix factorization (NMF), this type of analysis allowed us to accurately assign prey proteins to either the P-body or stress granule.

Besides defining organellar composition, PDB–MS can also identify contacts between organelles, as illustrated for example by the studies with APEX2 of contacts between the ER and the plasma membrane [35] or of those between the mitochondria and the ER [36]. Standard BioID was also recently used to identify proteins involved in ER-peroxisome contacts [37]. Taken together, these studies demonstrate the usefulness of PDB–MS for identifying the composition of intracellular structures, contacts between structures, or, in more general terms, the association of proteins that reside within 'insoluble' structures.

Signal amplification in proximity-dependent biotinylation

Notably, in addition to capturing weaker protein–protein interactions, PDB–MS can also identify 'cycling' proximity interactions, that is those that are context-dependent and/or characterized by rapid on-off rates, and which may be present at low abundance under steady-state conditions.

Using both AP–MS and BioID, we described a simple cycling interaction between a kinase (MST1) and a phospho-binding protein (MOB1). MST1 autophosphorylates, creating a docking site for MOB1. This phosphorylation event can be reversed by a PP2A-family phosphatase [38•] (Figure 4a). In normal growing cells, the active phosphatase maintains phosphorylated MST1 (and hence binding to MOB1) at low levels. AP–MS thus usually fails to detect this interaction unless phosphatase activity is abrogated. By contrast, BioID can detect the MST1–MOB1 interaction without phosphatase inhibition, which we attribute to an accumulation of cycling interactions. Since there is no de-biotinylase in human cells, biotinylation of a neighboring protein is permanent (as long as the protein is not degraded) and BioID labeling over several hours allows for the amplification of signal from these cycling interactions (Figure 4b). This concept is similar to the methylation-based M-Track approach developed to identify enzyme substrates in yeast [39]. Whether this phenomenon is at the root of other observations, for example the fact that proximity interactions can be observed between stress granule proteins in the absence of stress by both BioID and APEX [20•,31•], requires further study.

A similar approach was harnessed in using BioID to identify ubiquitin E3 ligase substrates. Coyaud *et al.* reasoned that the cycling (and relatively weak) interactions between E3 ligases and their substrates could be detected by BioID, providing that the targets were not subsequently degraded. By performing two parallel experiments—one in untreated cells and one in which degradation was inhibited with the proteasome inhibitor MG132—many of the known physiological targets of the SCF E3 ligase b-TrCP were recovered, and a number of new substrates were identified and validated [40•] (Figure 4c).

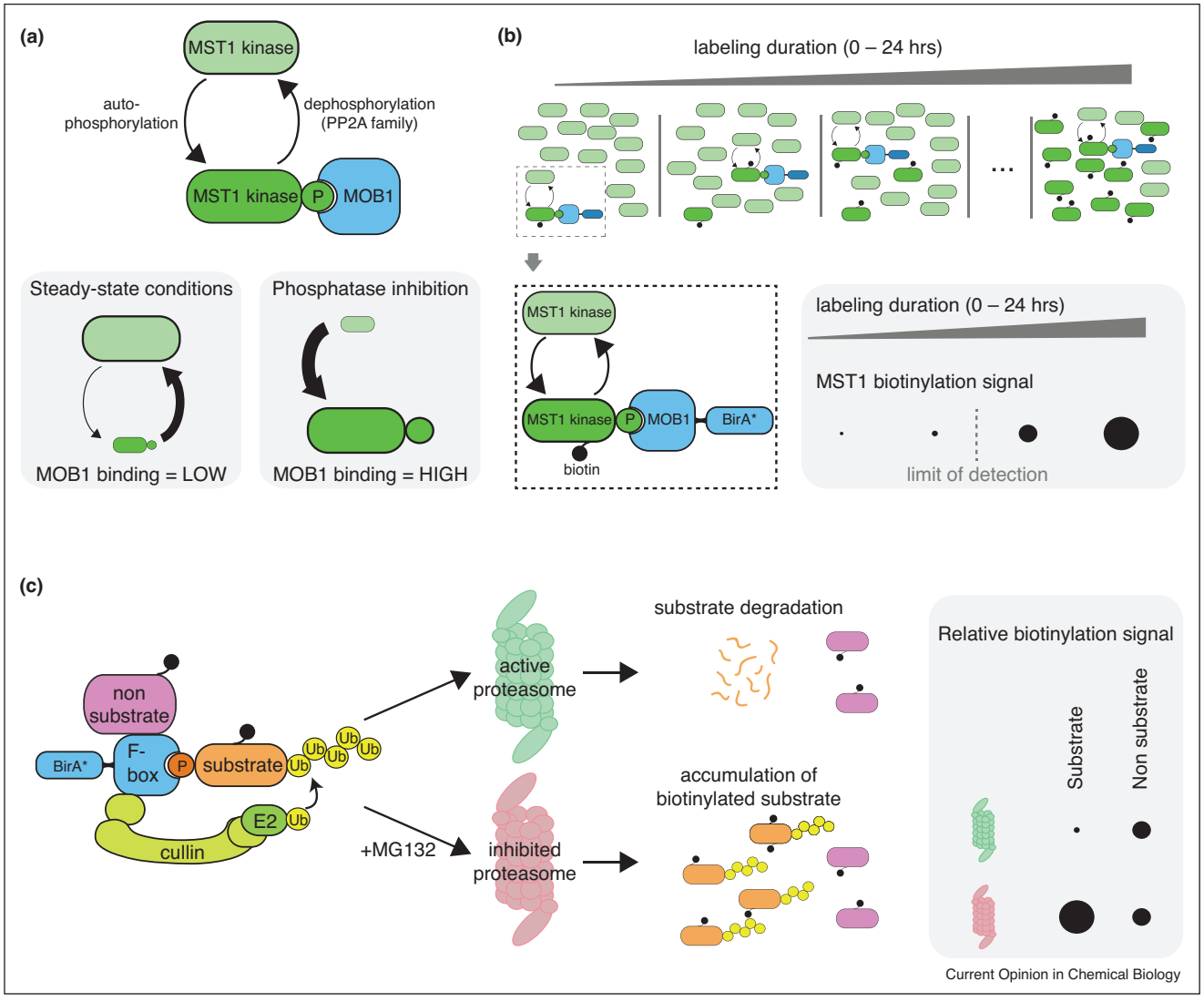
While in some cases, this 'signal amplification' is desirable (e.g. it can be harnessed to identify ubiquitin-ligase substrates), this aspect of PDB–MS can also render the interpretation of the resultant datasets more difficult, and can to some extent blunt the detection of regulated interactions, especially in the case of experiments with lengthy labeling periods such as those employing the original BioID tool. This also renders the validation of the interactions both more pressing and more challenging.

Experimental design and data analysis

Proximity-dependent biotinylation approaches performed on a single bait protein can generate hundreds to thousands of putative partner identifications (with parameters such as the amount of material and the sensitivity of the mass spectrometer significantly influencing these numbers). Importantly, however, in most cases the majority of these proteins are not true proximity partners for the bait of interest. Many are instead: biotin-labeled in the absence of the recombinant enzyme (e.g. endogenously biotinylated proteins such as mitochondrial carboxylases); proteins that are promiscuously biotinylated with most baits (e.g. with BioID in HEK293 cells, filamin A belongs to this category), or; proteins that bind non-specifically to the affinity support (sepharose or other bead types).

Controls should thus minimally include conditions that mimic endogenous biotinylation (such as no enzyme fused to the bait protein, or untransfected cells) and conditions that reproduce promiscuous biotinylation (e.g. enzyme alone expressed throughout the cell, and/or fused to an irrelevant polypeptide such as the Green Fluorescent Protein). To properly model promiscuous background, the control polypeptide must also be expressed to at least the same level as the bait protein, and should at least partially occupy the same intracellular locale. In cases where a protein of interest localizes to a specific subcellular location, it may also be useful to include an enzyme fusion that is specifically localized to this structure [41] (discussed further below). Notably, however, there are instances where controls can 'drown out' relevant signal. This can happen, for example, when an enzyme-fusion control protein is expressed at a much higher level than the bait under analysis. In these cases, reducing the expression level of the control protein to that

Figure 4



Signal amplification in PDB–MS and identification of enzyme–substrate interactions. **(a)** The kinase MST1 autophosphorylates to create a docking site for the MOB1 protein; a phosphatase reverses these phosphorylation and binding events. Under standard growth conditions, MST1 phosphorylation and MOB1 binding are low (usually below the detection limit of AP–MS). Inhibition of the phosphatase results in a marked increase in MST1 phosphorylation and in MOB1 binding. **(b)** In BioID, even though the fraction of autophosphorylated MST1 engaged in MOB1 binding remains low, the long labeling durations induce a build-up of biotinylation on the MST1 target. Eventually, this biotinylation signal passes the limit of detection, and is detected in the PDB–MS experiment. **(c)** Strategy for the identification of substrates for example F-box containing E3 ubiquitin ligases. F-box proteins interact with both non-substrate proteins (e.g. cullins, E2s and other stable binding partners) and substrates. Interaction with substrates promotes their polyubiquitylation and eventual degradation by the 26S proteasome. PDB–MS experiments conducted both in the presence and absence of a proteasome inhibitor (MG132) identifies proteins that display substrate behaviour (recovery is enhanced following proteasome inhibition) and non-substrates (levels are not significantly modulated in response to MG132 treatment).

of the bait should enable a more realistic estimation of background.

As with standard AP–MS approaches, multiple biological replicates of both bait and control experiments are critical to ensure that proximity interactions are robust and reproducible. The Contaminant Repository for Affinity Purifications [42] resource can be used to supplement controls for these types of analyses. Tools developed for the analysis of AP–MS datasets, including Significance Analysis of INTeractome (SAINT) [43], or CompPASS [44] can also be used to analyze proximity-dependent biotinylation data (see Ref. [45] for a review of scoring tools). While the selection of the tools is outside

the scope of this review, we further comment on the selection of controls specifically for PDB–MS in the next section.

Considerations and caveats for PDB–MS

The relative 'detectability' of proteins identified in a PDB–MS experiment depends not only on their distance from the bait, but factors such as the residence time of each protein in a complex (i.e. proteins that always reside together will be more efficiently labeled in theory than proteins that interact for shorter timespans), and the number of solvent-exposed reactive residues. Protein and complex size are additional confounding variables in this type of analysis. For example, PDB–MS conducted with a single component of a large protein complex may not label all components of the structure. Whether these results indicate that the bait is actually further from the unlabeled proteins in the complex than the proteins that are detected is not always clear. Nevertheless, this type of information can be informative; for example BioID profiles of dynein/dynactin complex proteins corroborated their structural organization [46], and approaches such as identifying and quantifying biotinylated peptides [47•,48,49•] can in theory help to infer structural organization.

For baits that are embedded in membranes or localized to organelles, abundant resident proteins (which are not necessarily direct interactors of the bait protein) often dominate the list of high-confidence identifications. For example, BioID of receptor tyrosine kinases identified hundreds of plasma membrane proteins ([50] e.g. 119 of the 167 hits with PTPRA were associated with GO:0005886: plasma membrane, p-value 6.6×10^{-29}). In these cases, the identification of direct interactors may require a different type of experimental design. In an ideal case, the control could be a related protein with the same localization, or a mutant version of the protein that loses specific interactions, but not localization. For example, PTPRA and PTPRF (a.k.a. phosphatase LAR) share many plasma membrane interactors [50], but PTPRA uniquely recovers its known interactor GRB2 [51] while PTPRF is strongly enriched in the liprin PPFIA1, as expected [52].

Alternatively, 'compartment controls' that provide a snapshot of the composition of a structure can be employed, for example by expressing an organelle marker or sequence tag (for instance, a nuclear localization sequence [41], farnesylation signal or mitochondrial targeting sequence [53]), with the caveat that these 'controls' may introduce their own bias. Rather than directly considering them as part of the scoring of contaminants, these compartment controls can also be used as a secondary 'enrichment' strategy to help define more specific proximity interactors. Lastly, for large datasets in which many baits are localized to the same structure, yet are expected to interact with different partners, one can analyze quantitative enrichment of a prey with a given bait across an entire dataset using for example the CompPASS tools developed by the Gygi and Harper groups for analyzing AP–MS data [44]. These tools, in addition to simpler enrichment analyses, are available both in the CRAPome [42] and through the ProHits-viz visualization toolbox [54]. As an ongoing project, we are generating a reference map of a human cell using BioID in HEK293 cells, and will make all data available on a web-based tool to help researchers interpret their own PDB–MS experiments.

A number of additional parameters could theoretically affect *in vivo* biotin labeling (and thus data interpretation). For example, the accessibility of the BirA*/APEX moiety in the context of a folded fusion protein could affect labeling, as the tag could be effectively 'buried' or fully exposed in the folded protein structure. The location of the tag on the protein (i.e. N-term versus C-term) can also be critically important in relation to membrane topology or orientation within a given structure. The position of the tagged polypeptide within a protein complex likely also plays an important role in labeling efficiency. For example, the protein of interest—along with its associated enzyme tag—could be exposed or completely buried within the context of a larger complex. Some bait protein localizations are also incompatible with biotin labeling (e.g. inside the lysosome). Finally, issues that are common to any tagging protocol are bait protein mislocalization caused by the tag (for example, due to disruption of a mitochondrial or peroxisomal targeting sequence) and/or disruption of protein–protein or protein–membrane interactions by the tag.

Since PDB–MS datasets will likely include a mixture of direct and indirect interactors, as well as proteins that simply reside in the same intracellular locale, how best to validate this type of data remains an open question. The selection of orthogonal approaches for additional characterization will be dependent on the localization, half-life and expression level of endogenous protein(s) of interest, amongst many other factors. We have predominantly relied on reciprocal PDB–MS and the acquisition of large datasets to overcome any overexpression artifacts and define specificity in associations [20•,27••,38•]. Standard biochemical approaches can certainly also be used, with the caveat that if the interactions do not withstand purification or involve proteins that are insoluble under standard purification conditions, they will likely be missed. Modifying buffer conditions or detergents as in Ref. [38•], or performing the validation experiment in the presence of a cross-linker, as in Ref. [55], can overcome some of these limitations. The use of non-biochemical validation assays such as the characterization of subcellular localization by immunofluorescence, or establishing co-localization through proximity ligation assays (PLA), can support

the results but these approaches are less quantitative and actually have a larger radius than PDB–MS. Lastly, techniques that are biased for direct interactors, including Protein Complementation Assays, Yeast Two-Hybrid or FRET-based approaches can certainly be used [56], but validation rates will be intrinsically low. While these orthogonal interaction/localization methods are certainly important, functional validation is the ultimate proof that the proximity-dependent interactions detected are meaningful. As an example, after identification of ~1,500 proteins in the centrosome/cilium PDB–MS map, we assessed the function of 500 of them by RNAi-mediated depletion followed by three cell-based assays, revealing distinct roles for ~300 of them [27••]. Integration of PDB–MS datasets with functional studies is therefore likely to further illuminate the biological processes under study.

Conclusion

Proximity-dependent biotinylation techniques are relatively simple to use, the data that they generate are highly complementary with more standard approaches, and these methods allow for the discovery of protein–protein interactions in previously inaccessible biological settings. As this family of techniques continues to evolve, we must ensure that our ability to properly analyse the data evolves with it. In this perspective, we have attempted to highlight some important points to keep in mind when considering proximity-dependent biotinylation approaches, including experimental design, data analysis and interpretation.

Conflict of interest statement

Nothing declared.

Acknowledgements

We are grateful to Ji-Young Youn for critical reading of this manuscript. We acknowledge funding from the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-139 to A.-C.G.), and the Canadian Institutes of Health Research (CIHR Foundation grant FDN143301 to A.-C.G. and CIHR MOP119289 and MOP130340 to B.R.). A.-C.G. is the Canada Research Chair in Functional Proteomics and the Lea Reichmann Chair in Cancer Proteomics. K.A. is a recipient of CGS(M) studentship from CIHR.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Yates JR 3rd, Gilchrist A, Howell KE, Bergeron JJ: **Proteomics of organelles and large cellular structures**. *Nat Rev Mol Cell Biol* 2005, **6**:702-714.

2. Christoforou A, Mulvey CM, Breckels LM, Geladaki A, Hurrell T, Hayward PC, Naake T, Gatto L, Viner R, Martinez Arias A *et al.*: **A draft map of the mouse pluripotent stem cell spatial proteome**. *Nat Commun* 2016, **7**:8992.

3. Itzhak DN, Tyanova S, Cox J, Borner GH: **Global, quantitative and dynamic mapping of protein subcellular localization**. *eLife* 2016, **5**.

4. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S *et al.*: **A census of human soluble protein complexes**. *Cell* 2012, **150**:1068-1081.

5. Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A *et al.*: **Panorama of ancient metazoan macromolecular complexes**. *Nature* 2015, **525**:339-344.

6. Kristensen AR, Gsponer J, Foster LJ: **A high-throughput approach for measuring temporal changes in the interactome**. *Nat Methods* 2012, **9**:907-909.

7. Jean Beltran PM, Mathias RA, Cristea IM: **A portrait of the human organelle proteome in space and time during cytomegalovirus infection**. *Cell Syst* 2016, **3**:361-373 e366.

8. Gingras AC, Raught B: **Beyond hairballs: the use of quantitative mass spectrometry data to understand protein-protein interactions**. *FEBS Lett* 2012, **586**:2723-2731.

9. Nooren IM, Thornton JM: **Diversity of protein-protein interactions**. *EMBO J* 2003, **22**:3486-3492.

10. Wang X, Huang L: **Identifying dynamic interactors of protein complexes by quantitative mass spectrometry**. *Mol Cell Proteomics* 2008, **7**:46-57.

11. •• Roux KJ, Kim DI, Raida M, Burke B: **A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells**. *J Cell Biol* 2012, **196**:801-810.
First description of the use of a biotin ligase to mediate covalent labeling of proximity interactors that can be captured and identified by mass spectrometry. The approach, called BioID by the authors, was used to characterize the nuclear lamina.

12. Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Roux KJ: **An improved smaller biotin ligase for BioID proximity labeling**. *Mol Biol Cell* 2016, **27**:1188-1196.

13. Ramanathan M, Majzoub K, Rao DS, Neela PH, Zarnegar BJ, Mondal S, Roth JG, Gai H, Kovalski JR, Siprashvili Z *et al.*: **RNA-protein interaction detection in living cells**. *Nat Methods* 2018, **15**:207-212.

14. • Branon T: **Efficient proximity labeling in living cells and organisms with TurboID**. *Nature Biotechnol* 2018, **36**:880-887.
Directed evolution of biotin ligase to improve its activity, with the resulting proteins capable of mediating biotinylation of prey proteins in as little as 10 min. The evolved enzymes, called TurboID and miniTurbo, were also benchmarked in whole organisms, namely*Drosophila* and *C. elegans*

15. •• Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY: **Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging**. *Science* 2013, **339**:1328-1331.
First use of a peroxidase to perform PDB–MS, and use of the fusion enzyme coupled to quantitative proteomics to define the composition of the mitochondrial matrix without the need for mitochondrial purification.

16. Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY: **Directed evolution of APEX2 for electron microscopy and proximity labeling**. *Nat Methods* 2015, **12**:51-54.

17. •• Loh KH, Stawski PS, Draycott AS, Udeshi ND, Lehrman EK, Wilton DK, Svinkina T, Deerinck TJ, Ellisman MH, Stevens B *et al.*: **Proteomic analysis of unbounded cellular compartments: synaptic clefts**. *Cell* 2016, **166**:1295-1307 e1221.
Demonstration of PDB–MS using horseradish peroxidase in an extracellular compartment, the synaptic cleft.

18. Jiang S, Kotani N, Ohnishi T, Miyagawa-Yamguchi A, Tsuda M, Yamashita R, Ishiura Y, Honke K: **A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction**. *Proteomics* 2012, **12**:54-62.

19. Li XW, Rees JS, Xue P, Zhang H, Hamaia SW, Sanderson B, Funk PE, Farndale RW, Lilley KS, Perrett S *et al.*: **New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay**. *J Biol Chem* 2014, **289**:14434-14447.

20. • Youn JY, Dunham WH, Hong SJ, Knight JDR, Bashkurov M, Chen GI, Bagci H, Rathod B, MacLeod G, Eng SWM *et al.*: **High-density proximity mapping reveals the subcellular**

organization of mRNA-associated granules and bodies. *Mol Cell* 2018, **69**:517-532 e511.
Largest study to date (in number of individual baits) employing PDB–MS. The experimental design and scope of the experiments enabled the development of alternative ways to explore PDB–MS data to reconstruct the organization of membrane-less organelles.

21. • Lobingier BT, Huttenhain R, Eichel K, Miller KB, Ting AY, von Zastrow M, Krogan NJ: **An approach to spatiotemporally resolve protein interaction networks in living cells**. *Cell* 2017, **169**:350-360 e312.
Demonstration of the usefulness of PDB–MS approaches for the identification of rapidly regulated interactions in the context of G-protein-coupled receptor signaling; also see Ref. [22•].

22. • Paek J, Kalocsay M, Staus DP, Wingler L, Pascolutti R, Paulo JA, Gygi SP, Kruse AC: **Multidimensional tracking of GPCR signaling via peroxidase-catalyzed proximity labeling**. *Cell* 2017, **169**:338-349 e311.
Demonstration of the usefulness of PDB–MS approaches for the identification of rapidly regulated interactions in the context of G-protein-coupled receptor signaling; also see Ref. [21].

23. Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, Roux KJ: **Probing nuclear pore complex architecture with proximity-dependent biotinylation**. *Proc Natl Acad Sci U S A* 2014, **111**:E2453-E2461.

24. Milo R, Philips R: *Cell biology by the numbers*. New York, NY: Garland Science, Taylor & Francis Group; 2016.

25. Yang Z, Jakymiw A, Wood MR, Eystathioy T, Rubin RL, Fritzler MJ, Chan EK: **GW182 is critical for the stability of GW bodies expressed during the cell cycle and cell proliferation**. *J Cell Sci* 2004, **117**:5567-5578.

26. Erickson HP: **Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy**. *Biol Proced Online* 2009, **11**:32-51.

27. •• Gupta GD, Coyaud E, Goncalves J, Mojarad BA, Liu Y, Wu Q, Gheiratmand L, Comartin D, Tkach JM, Cheung SW *et al.*: **A dynamic protein interaction landscape of the human centrosome-cilium interface**. *Cell* 2015, **163**:1484-1499.
First comprehensive use of PDB–MS to map the composition and structural organization of a membrane-less organelle, the centrosome.

28. Xie W, Chojnowski A, Boudier T, Lim JS, Ahmed S, Ser Z, Stewart C, Burke B: **A-type lamins form distinct filamentous networks with differential nuclear pore complex associations**. *Curr Biol* 2016, **26**:2651-2658.

29. Dong JM, Tay FP, Swa HL, Gunaratne J, Leung T, Burke B, Manser E: **Proximity biotinylation provides insight into the molecular composition of focal adhesions at the nanometer scale**. *Sci Signal* 2016, **9** rs4.

30. Guo Z, Neilson LJ, Zhong H, Murray PS, Zanivan S, Zaidel-Bar R: **E-cadherin interactome complexity and robustness resolved by quantitative proteomics**. *Sci Signal* 2014, **7** rs7.

31. • Markmiller S, Soltanieh S, Server KL, Mak R, Jin W, Fang MY, Luo EC, Krach F, Yang D, Sen A *et al.*: **Context-dependent and disease-specific diversity in protein interactions within stress granules**. *Cell* 2018, **172**:590-604 e513.
Use of APEX to profile the composition of stress granules across cell types and disease states.

32. Firat-Karalar EN, Rauniyar N, Yates JR 3rd, Stearns T: **Proximity interactions among centrosome components identify regulators of centriole duplication**. *Curr Biol* 2014, **24**:664-670.

33. Liu X, Salokas K, Tamene F, Jiu Y, Weldatsadik RG, Ohman T, Varjosalo M: **An AP–MS—and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations**. *Nat Commun* 2018, **9**:1188.

34. Bersuker K, Peterson CWH, To M, Sahl SJ, Savikhin V, Grossman EA, Nomura DK, Olzmann JA: **A proximity labeling strategy provides insights into the composition and dynamics of lipid droplet proteomes**. *Dev Cell* 2018, **44**:97-112 e117.

35. Jing J, He L, Sun A, Quintana A, Ding Y, Ma G, Tan P, Liang X, Zheng X, Chen L *et al.*: **Proteomic mapping of ER-PM junctions identifies STIMATE as a regulator of Ca(2)(+) influx**. *Nat Cell Biol* 2015, **17**:1339-1347.

36. Hung V, Lam SS, Udeshi ND, Svinkina T, Guzman G, Mootha VK, Carr SA, Ting AY: **Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation**. *eLife* 2017, **6**.

37. Hua R, Cheng D, Coyaud E, Freeman S, Di Pietro E, Wang Y, Vissa A, Yip CM, Fairn GD, Braverman N *et al.*: **VAPs and ACBD5 tether peroxisomes to the ER for peroxisome maintenance and lipid homeostasis**. *J Cell Biol* 2017, **216**:367-377.

38. • Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, Lin ZY, Bagshaw RD, Sicheri F, Pawson T *et al.*: **Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions**. *Sci Signal* 2013, **6** rs15.
Early parallel analysis of signaling components using AP–MS and PDB–MS. These comparative analyses were instrumental in understanding the key differences between these two types of approaches to identify protein–protein interactions.

39. Zuzuarregui A, Kupka T, Bhatt B, Dohnal I, Mudrak I, Friedmann C, Schuchner S, Frohner IE, Ammerer G, Ogris E: **M-Track: detecting short-lived protein-protein interactions in vivo**. *Nat Methods* 2012, **9**:594-596.

40. • Coyaud E, Mis M, Laurent EM, Dunham WH, Couzens AL, Robitaille M, Gingras AC, Angers S, Raught B: **BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates**. *Mol Cell Proteomics* 2015, **14**:1781-1795.
Demonstration that PDB–MS can be used to identify enzyme substrates for E3 ubiquitin ligases.

41. Lambert JP, Tucholska M, Go C, Knight JD, Gingras AC: **Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes**. *J Proteomics* 2015, **118**:81-94.

42. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY *et al.*: **The CRAPome: a contaminant repository for affinity purification-mass spectrometry data**. *Nat Methods* 2013, **10**:730-736.

43. Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI: **SAINT: probabilistic scoring of affinity purification-mass spectrometry data**. *Nat Methods* 2011, **8**:70-73.

44. Sowa ME, Bennett EJ, Gygi SP, Harper JW: **Defining the human deubiquitinating enzyme interaction landscape**. *Cell* 2009, **138**:389-403.

45. Armean IM, Lilley KS, Trotter MW: **Popular computational methods to assess multiprotein complexes derived from label-free affinity purification and mass spectrometry (AP–MS) experiments**. *Mol Cell Proteomics* 2013, **12**:1-13.

46. Redwine WB, DeSantis ME, Hollyer I, Htet ZM, Tran PT, Swanson SK, Florens L, Washburn MP, Reck-Peterson SL: **The human cytoplasmic dynein interactome reveals novel activators of motility**. *eLife* 2017, **6**.

47. • Udeshi ND, Pedram K, Svinkina T, Fereshetian S, Myers SA, Aygun O, Krug K, Clauser K, Ryan D, Ast T *et al.*: **Antibodies to biotin enable large-scale detection of biotinylation sites on proteins**. *Nat Methods* 2017, **14**:1167-1170.
Alternative workflow to streptavidin capture for the identification of biotinylated peptides in PDB–MS experiments.

48. Kim DI, Cutler JA, Na CH, Reckel S, Renuse S, Madugundu AK, Tahir R, Goldschmidt HL, Reddy KL, Huganir RL *et al.*: **BioSITe: a method for direct detection and quantitation of site-specific biotinylation**. *J Proteome Res* 2018, **17**:759-769.

49. • Mackmull MT, Klaus B, Heinze I, Chokkalingam M, Beyer A, Russell RB, Ori A, Beck M: **Landscape of nuclear transport receptor cargo specificity**. *Mol Syst Biol* 2017, **13**:962.
Systematic use of PDB–MS to identify the cargo specificity of nuclear transport receptor; identification of the biotinylated sites assisted in defining interaction mechanisms.

50. St-Denis N, Gupta GD, Lin ZY, Gonzalez-Badillo B, Veri AO, Knight JDR, Rajendran D, Couzens AL, Currie KW, Tkach JM *et al.*: **Phenotypic and interaction profiling of the human phosphatases identifies diverse mitotic regulators**. *Cell Rep* 2016, **17**:2488-2501.

51. Su J, Batzer A, Sap J: **Receptor tyrosine phosphatase R-PTP-alpha is tyrosine-phosphorylated and associated with the adaptor protein Grb2**. *J Biol Chem* 1994, **269**:18731-18734.

52. Serra-Pages C, Medley QG, Tang M, Hart A, Streuli M: **Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins**. *J Biol Chem* 1998, **273**:15611-15620.

53. Cole A, Wang Z, Coyaud E, Voisin V, Gronda M, Jitkova Y, Mattson R, Hurren R, Babovic S, Maclean N *et al.*: **Inhibition of the mitochondrial protease ClpP as a therapeutic strategy for human acute myeloid leukemia**. *Cancer Cell* 2015, **27**:864-876.

54. Knight JDR, Choi H, Gupta GD, Pelletier L, Raught B, Nesvizhskii AI, Gingras AC: **ProHits-viz: a suite of web tools for visualizing interaction proteomics data**. *Nat Methods* 2017, **14**:645-646.

55. Subbotin RI, Chait BT: **A pipeline for determining protein-protein interactions and proximities in the cellular milieu**. *Mol Cell Proteomics* 2014, **13**:2824-2835.

56. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I: **Fundamentals of protein interaction network mapping**. *Mol Syst Biol* 2015, **11**:848.