Research Techniques Made Simple: Emerging Methods to Elucidate Protein Interactions through Spatial Proximity

Yonglu Che¹ and Paul A. Khavari¹,²

Interaction between proteins is essential for fundamental cellular processes, and the diversity of such interactions enables the vast variety of functions essential for life. A persistent goal in biological research is to develop assays that can faithfully capture different types of protein interactions to allow their study. A major step forward in this direction came with a family of methods that delineates spatial proximity of proteins as an indirect measure of protein-protein interaction. A variety of enzyme- and DNA ligation-based methods measure protein co-localization in space, capturing novel interactions that were previously too transient or low affinity to be identified. Here we review some of the methods that have been successfully used to measure spatially proximal protein-protein interactions.


Interactions between proteins are essential for fundamental cellular processes, and the diversity of such interactions enables the vast variety of functions essential for life. A persistent goal in biological research is to develop assays that can faithfully capture different types of protein interactions to allow their study. A major step forward in this direction came with a family of methods that delineates spatial proximity of proteins as an indirect measure of protein-protein interaction. A variety of enzyme- and DNA ligation-based methods measure protein co-localization in space, capturing novel interactions that were previously too transient or low affinity to be identified. Here we review some of the methods that have been successfully used to measure spatially proximal protein-protein interactions.

CME Activity Dates: November 20, 2017
Expiration Date: November 19, 2018
Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: All authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationships(s) with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:

- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of Beaumont Health and the Society for Investigative Dermatology. Beaumont Health is accredited by the ACCME to provide continuing medical education for physicians. Beaumont Health designates this enduring material for a maximum of 1.0 AMA PRA Category 1 Credit(s)™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the Journal of Investigative Dermatology website: http://www.jidonline.org/current. Tests for CME credits may only be submitted online at https://cme@beaumont.cloud-cme.com/RTMS-Dec17 – click ‘CME on Demand’ and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

Interactions between proteins underlie a significant amount of the mechanical, structural, and signaling processes that are necessary to support various functions of living cells. The variety of protein-protein interactions is highly diverse and heavily context dependent. Methods to study protein-protein interactions like affinity-capture complex purification (LaCava et al., 2016), surface plasmon resonance (Schuck, 1997), isothermal titration calorimetry (Velazquez-Campoy, 2006), yeast-two-hybrid screening (Miller et al., 2004), and fluorescence resonance energy transfer (Heim et al., 1996) among many others have provided key insights to understanding function, but each method suffers from limitations. Most existing methods require stable protein interactions to survive the harsh processing steps necessary to extract proteins from cells, and many others query interactions outside the native cellular context, relying on empirically reconstituted conditions (Table 1).

Proximity-based protein labeling attempts to address some of these limitations through a different approach. Instead of purifying protein complexes or searching for evidence of
interactions between recombinant proteins in vitro, proximity labeling allows identification of proteins that reside within a 10- to 20-nm radius of an introduced labeling enzyme. Such a labeling strategy allows surveillance of transient and weak interactions. Because proximity labeling can be done in living cells, it can also detect interactions that require fragile macromolecular assemblies, intact subcellular structures, lipid or nucleic acid cofactors, and posttranslational modifications that are difficult to retain or reconstitute in vitro. This idea of using spatial information to expand and inform the networks of interactions inside a cell is increasingly being applied in new techniques and applications. This review focuses on how proximity labeling has been successfully used thus far, the limitations of the data it provides, and the potential for further development.

**ENZYMATIC PROXIMITY LABELING (BirA, APEX, HRP)**

At the heart of proximity labeling are enzymes that produce distance-constrained reactive biotin molecules. Although spatial proximity of proteins within a cell could to some extent be measured by modern microscopy techniques, the reagents and time necessary make microscopy impractical as a high-throughput discovery method. In contrast, labeling followed by protein isolation allows for unbiased discovery through mass spectrometry. The discovery of several enzymes that generate short-lived, reactive biotin intermediates, providing the mechanism by which spatial labeling is possible.

Proximity labeling can also be used to label cellular compartments using localization sequences or to perform conditional proteomics with split versions of the enzymes.

DNA-based proximity ligation is an alternative, low-throughput method that also assays spatial proximity.

**SUMMARY**

- Proximity protein labeling allows detection of spatially proximal proteins.
- BirA, Apex, and HRP are enzymes that generate spatially confined reactive biotin intermediates, providing the mechanism by which spatial labeling is possible.
- Proximity labeling can also be used to label cellular compartments using localization sequences or to perform conditional proteomics with split versions of the enzymes.
- DNA-based proximity ligation is an alternative, low-throughput method that also assays spatial proximity.

Biotin intermediates have made them useful for proximity labeling. HRP and ascorbate peroxidase are both peroxidases that use hydrogen peroxide to perform oxidation reactions. HRP has been a staple enzyme in activity assays, finding wide applications from ELISAs to Western blotting long before its application in proximity labeling. HRP is unfortunately inactive in the cytosol, necessitating investigation of other peroxidases. Ascorbate peroxidase was therefore engineered as an HRP substitute and given the new name APEX (Martell et al., 2012). APEX was first applied to electron microscopy but quickly found a role in proximity-dependent protein labeling because of its ability to rapidly generate short-lived, spatially confined reactive biotin intermediates.

When a construct of these labeling enzymes fused with a protein of interest is expressed within a cell, the fusion protein can properly localize and perform its usual biological functions (Figure 1a–c). Proteins that stay within the labeling radius of the enzyme longer than would be expected by random motion become enriched in the total biotinylated subset of proteins. During this time, cells can be exposed to biochemical and genetic perturbations appropriate for the given experiment. At the end of the biotin-labeling period, cells can be lysed to isolate total protein (Figure 1d). Because biotin labels are covalently linked to proteins themselves, lysis conditions are unlikely to introduce artifacts, providing a significant advantage to affinity purification methods where lysis conditions must be carefully chosen to preserve interactions and where lysis itself can compromise the separation of cellular compartments and lead to false positive and false negative findings (Table 2). A key advantage of the biotin handle is biotin’s strong but reversible affinity for streptavidin (Chivers et al., 2011). The high affinity of the biotin-streptavidin complex allows biotinylated proteins to be efficiently captured from the lysed solution and then stringently washed to remove nonspecific interactions (Figure 1e). After protein elution and digestion, samples can be subjected to Western blotting to query specific interactions or mass spectrometry to broadly map all of the spatially proximal proteins detected (Figure 1f). The fidelity of resulting candidate interactions can be tested by determining if previously known protein interactions were detected and by using gene ontology annotations to cross-check for functions known to be associated with the protein of interest (Figure 1g). Novel associations between proteins can frequently be shown in such data. However, the veracity of each detected association invariably requires further validation by orthogonal technical approaches. An example of how proximity proteomics can be used to investigate the novel interactors is illustrated by Perez-White et al. (2017) in their study of the receptor tyrosine kinase EphA2 (Figure 1h and i).

The choice of labeling enzyme must be carefully considered for a number of practical experimental reasons. BirA converts biotin into reactive biotin—adenosine monophosphate, whereas APEX and HRP create free radicals of biotin as intermediates. BirA can be used in living tissue, as has been successfully applied within organotypic skin models (Perez-White et al., 2017). Unlike the simple biotin used for BirA, the biotin-phenol reagent used for APEX and the hydrogen peroxide reagent used for HRP are not suitable for organoid or in vivo studies because of the tissue toxicity of
Table 1. Comparison of techniques to study protein-protein interactions

<table>
<thead>
<tr>
<th>Technique</th>
<th>Spatial Proximity or Direct Interaction</th>
<th>Maintains Cellular Context</th>
<th>High/Low Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximity-dependent protein labeling</td>
<td>Spatial proximity</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Proximity-dependent DNA ligation</td>
<td>Spatial proximity</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>Affinity-capture complex purification</td>
<td>Direct interaction</td>
<td>Yes²</td>
<td>High</td>
</tr>
<tr>
<td>Surface plasmon resonance</td>
<td>Direct interaction</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Isothermal titration calorimetry</td>
<td>Direct interaction</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Yeast-two-hybrid screening</td>
<td>Direct interaction</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Florescence resonance energy transfer</td>
<td>Spatial proximity</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

1A brief list of representative techniques used to study protein-protein interactions. Techniques that measure spatial proximity also detect direct interaction because direct interaction is usually within the labeling radius. Techniques categorized as high throughput can be adapted as low-throughput techniques but not vice versa.

2Although protein complexes form within the proper cellular context, artifacts can arise during cell lysis.

Figure 1. Proximal protein labeling workflow. (a) A protein of interest for which proximal proteins are to be interrogated is first selected and fused on a selected terminus to an enzyme to catalyze biotinylation. (b) A functional fusion protein expressed at physiologic levels in the cells interacts with normal protein partners. (c) Stimulation of the biotinylation enzyme covalently labels proteins in an approximately $1 \times 10^8$-m space around the fusion protein. (d) Denaturing cell lysis is used to liberate all proteins and (e) biotinylated proteins are isolated by streptavidin pulldown. (f) After elution off of streptavidin beads, the biotinylated proteins can be detected in a sensitive but low-throughput manner in Western blotting or more comprehensively by mass spectrometry. (g) Comparison of proximal protein interactors when analyzed with proper control datasets can capture both previously known interactions and potentially novel ones. (h) Perez White et al. (2017) used a BirA tag to identify the proximal protein interactome of the receptor tyrosine kinase EphA2 in primary human keratinocytes and reconstituted three-dimensional human epithelium. The broad landscape of the interactome can be partially understood by identifying common pathway annotations of EphA2 interactors. (i) Individual interacting proteins of interest are visualized by peptide spectral matches. The BirA-tagged protein is commonly among the most highly enriched because of self-biotinylation but in the case of EphA2 could also be caused by receptor dimerization.
both phenol and peroxide-containing reagents. APEX and HRP can label significant amounts of protein in minutes, but BirA needs hours to a day for optimal labeling. APEX and HRP may thus be more suitable for experiments where temporal resolution is critical, whereas BirA is more suitable for studying interactions during homeostasis. The amino acids favored for reaction also differ by enzyme. BirA favors lysine residues, whereas APEX and HRP prefer tyrosine, so proteins detected by one method may not necessarily be detected by the other because of differences in accessibility of target residues. In addition, the cell’s many compartments can have drastically different conditions, changing how each protein and the reactive biotin intermediates behave. HRP, for example, contains four disulfide links critical to its structure. It is therefore an impractical choice for the study of proteins primarily localized to the reducing environment of the cytoplasm, but it functions well on the cell surface or in the endoplasmic reticulum and Golgi apparatus. Reactive intermediates also are likely to behave slightly differently depending on where they are generated in the cell, underscoring the need for corroboration of proximity proteomics findings using orthogonal methods.

Proximity-dependent proteomics have been successfully applied a number of times to study the interactions of proteins in the epidermis. EphA2, a receptor tyrosine kinase and regulator of epidermal homeostasis, was discovered to facilitate tight junction formation through interaction with afadin (Perez-White et al., 2017). In the process of epidermal differentiation, proximity-dependent labeling led to the discovery of interactions between MPZL3 and FDXR (Bhaduri et al., 2015), as well as CALM5 and SFN (Sun et al., 2015). MPZL3-FDXR interaction drives an increase in reactive oxygen species to progress differentiation, and CALM5-SFN interaction controls expression of late differentiation genes.

CAVEATS AND CONTROLS

Although proximal protein labeling shows previously undetected interactions and can help prioritize exploration into unknown protein functions, it is important to remember that the reliability of the method strongly correlates with the fidelity with which the protein-enzyme fusion behaves like the native protein. When fusion of the selected labeling enzyme compromises function or localization of the linked protein of interest, or when exogenous expression delivers supraphysiologic levels of protein, the interactome will be less faithful to that of the endogenous protein. Additionally, the linker between enzyme and protein of interest must be carefully considered. Both the length and composition of the linking sequence will determine the range and accessibility of the enzyme to target proteins (Chen et al., 2013). Another important caveat when interpreting such data is that proximal protein labeling shows only spatial co-localization. Determining whether protein co-localization occurs because of direct or indirect protein-protein interactions requires separate analysis by other methodologies.

As with other high-throughput methods for discovering protein-protein interactions, proper controls must be included to determine the likelihood that a discovered interaction is biologically meaningful. Although the most appropriate controls must be determined for each individual experiment,
a generally useful control is to express the biotinylating enzyme alone. Biotinylated proteins captured from cells expressing the biotinylating enzyme in the absence of a protein of interest include proteins that are biotinylated as part of their usual function, highly expressed proteins that are likely to be background, and proteins with affinity to the labeling enzyme itself. Proteins that are robustly represented in this control dataset are likely to be background and are often excluded from further analysis. More specific controls include mutants of the protein of interest that result in specific loss or gain of function or localization sequences to enrich for the background labeling in specific cellular compartments in which a protein of interest is known to reside.

APPLICATIONS OF PROXIMAL PROTEIN LABELING

Proximity labeling has been used most commonly to study protein-protein interactions, but several additional uses of the labeling enzymes are worth mentioning.

Cellular compartment labeling

One major caveat to proximity labeling is that the method reflects only spatial co-localization, not direct interaction. Hung et al. (2016) have used this to their advantage to study the protein populations that make up different compartments of the mitochondria. By linking APEX to localization tags targeted to the cytoplasm, the mitochondrial intermembrane space, and the inner membrane space, proteins uniquely expressed in each compartment have been identified (Figure 2a).

Cell surface labeling

Studying proteins expressed at the cell surface poses challenges and opportunities different from those of an intracellular protein. An opportunity to study an unaltered, endogenous protein arises if an antibody or complementary small molecule exists to the protein of interest and can be fused to HRP (Miyagawa-Yamaguchi et al., 2014). A complementary molecule-HRP fusion labels proteins proximal to the endogenous protein of interest on the cell surface without having to introduce an exogenous construct (Figure 2b).

Time-resolved agonist response

The short (~1-minute) labeling period of APEX allows time course studies to be performed for rapid biological processes. Paek et al. (2017) fused APEX to G protein-coupled receptors (angiotensin II type 1 receptor and the β2 adrenoceptor) to study the dynamic receptor-agonist interactions (Paek et al., 2017). Lobinger et al. (2017) took a similar approach, also with G protein-coupled receptors (B2AR and DOR). In cases like this, proximity labeling can further elucidate previously known interactions by permitting facile measurement of such interactions through time (Figure 2c).

Conditional proteomics with split proteins

As it is becoming increasingly clear that proteins interact within vast and intricate networks and complexes, the ability to perform conditional proteomics is a rising challenge. BirA has been adapted for such a purpose by splitting the protein into two nonfunctioning parts that regain function when localized together (Schopp et al., 2017). By fusing the two halves to different proteins, proximity labeling can be limited to only the occasions during which those two proteins interact (Figure 2d). This type of conditional labeling may be a tool with which puzzles of protein complex assembly can be solved with minimal starting information about the constituents of such complexes.

VALIDATING PROXIMAL INTERACTION THROUGH DNA LIGATION

Many methods exist to test protein-protein interactions, prominent among which is co-immunoprecipitation, but there is a relative shortage of methods outside of proximity labeling to study protein proximity. Microscopy and fluorescence resonance energy transfer are useful tools, but proximity ligation analysis was developed as an easier method to
address spatial proximity of two endogenous proteins (Fredriksson et al., 2002). In fixed and permeabilized cells, a binary measure of whether or not two proteins reside within 30–40 nm of each other can be achieved by targeting antibodies to each of the two proteins. Attached to these antibodies are two oligonucleotides that can facilitate the ligation of a free-floating connector probe (Figure 3a). The two oligonucleotides must both be present for successful ligation, and successful ligation creates the template for an isothermal rolling circle amplification reaction. This reaction can be performed to incorporate fluorescent dye, resulting in a local hotspot of fluorescence whenever the two target proteins are within close vicinity to each other. Over an entire cell or tissue, detected interactions can be visualized microscopically by fluorescent clusters, each representing one instance of protein-protein co-localization. Perez White et al. (2017) use this technique to validate an interaction initially discovered with proximity labeling of EpHA2 (Figure 3b).

FUTURE DIRECTIONS IN PROXIMAL PROTEIN LABELING

The principles underlying use of proximity labeling by BirA, APEX, and HRP came from an earlier technique, DNA adenine methyltransferase identification (i.e., DamID) that maps DNA-protein interactions using a spatially confined methyltransferase reaction (Van Steensel et al., 2000). True to its origins, proximal protein labeling has been adapted back to studying DNA-protein interactions. Although DNA adenine methyltransferase identification allows one to pick a protein of interest and investigate proximal DNA sequences, CRISPR/Cas9-APEX does the reverse: first identifying a genomic locus of interest and then identifying proteins in the immediate vicinity (Myers et al., 2017). Improvements of methods to study DNA-protein interactions with proximity labeling and its adaptation to study protein-RNA interactions are likely to soon follow. Although it is not currently possible, the study of proximal protein interactomes in living organisms would be an invaluable insight into tissue-specific protein biology. With the development of more efficient labeling enzymes and given the robustness of the biotin-streptavidin interaction, proximity labeling in vivo may soon be achievable.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the USVA Office of Research and Development and by NIAMS/NIH AR043799 to PAK. YC is a recipient of Medical Scientist Training Program (MSTP) support from Stanford University.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the paper. Teaching slides are available as supplementary material.

REFERENCES


