The evolving capabilities of enzyme-mediated proximity labeling
Ying Zhou¹ and Peng Zou¹,²

Abstract
The subcellular organization of proteins and RNA molecules is crucial for their proper functions. Over the past decade, both ligase-mediated and peroxidase-mediated proximity labeling (PL) techniques have been developed to map biomolecules at near-nanometer spatial resolution and subminute temporal resolution. These methods are shedding light on the spatial arrangement of proteome and transcriptome in their native context. Here, we review the recent evolution and applications of PL techniques, compare and contrast the two classes of methods, and highlight emerging trends and future opportunities.

Addresses
¹ College of Chemistry and Molecular Engineering, Synthetic and Functional Biomolecules Center, Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, Peking University, Beijing, 100871, China
² Peking-Tsinghua Center for Life Sciences, PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China

Corresponding author: Zou, Peng (zoupeng@pku.edu.cn)

Current Opinion in Chemical Biology 2021, 60:30–38
This review comes from a themed issue on Omics
Edited by Nicholas Scott and Laura Edgington-Mitchell
For a complete overview see the Issue and the Editorial
https://doi.org/10.1016/j.cbpa.2020.06.013
1367-5931/© 2020 Elsevier Ltd. All rights reserved.

Keywords
Proximity labeling, APEX, BioID, Peroxidase, Biotin ligase.

Introduction
Eukaryotic cells are elaborately divided into subcellular compartments that feature distinct biochemical characteristics. The spatial organization of proteins and RNAs in these subcellular regions is intimately linked to their biological functions, including signal transduction [1,2], localized protein synthesis [3,4], regulation of chromatin structure [5], etc. While most famously observed at membrane-bound organelles (either in the interior [6] or on the surface [7]), the subcellular targeting of proteins and RNAs have also been discovered in membrane-less compartments, such as highly dynamic liquid-like condensates that form via liquid–liquid phase separation (LLPS) [8]. For example, stalled translation-initiation causes mRNAs and mRNA-binding proteins (RBPs) to assemble into stress granules (SGs), in a process of protecting cells from oxidative damages or other cellular stress [9].

The subcellular proteome and transcriptome have been traditionally investigated by co-immunoprecipitation (co-IP) and biochemical fractionation. However, both methods require prior cell lysis, which is prone to losing low-affinity and transient protein–protein interactions. In addition, co-IP is limited by the availability of high-quality antibodies against the bait, while biochemical fractionation often suffers from incomplete purification. For example, the transcriptomic profiling of isolated mitochondria has identified abundant contaminations from the cytoplasm [10]. Furthermore, not all subcellular structures are amenable to fractionation [11]. Over the past decade, enzyme-mediated proximity labeling (PL) techniques have emerged as powerful tools for locating proteins and RNAs in live cells. In these methods, an engineered enzyme is expressed at a specific subcellular locale, where it catalyzes the in situ synthesis of a highly reactive small-molecule intermediate, which subsequently diffuses away and reacts with proteins and/or RNAs to form a covalent label (Figure 1a). Due to its limited lifetime, the local density, and hence the labeling efficiency of the intermediate drops off as a function of the distance from the enzyme. Thus, all else being equal, proteins/RNAs proximal to the enzyme are more likely to be labeled than distal ones. Compared with biochemical fractionation, PL could access information from subcellular compartments that are impossible to purify or highly dynamic (e.g. signaling complexes, LLPS, etc.). PL also complements co-IP studies because it is capable of mapping distant protein–protein interactions, with an ‘action contour map’ that spans over several ‘interaction layers’ (Figure 1a).

In this review, we highlight several emerging trends of PL technology development and new avenues of its applications. As discussed below, enzyme-mediated PL is now moving rapidly from membrane-enclosed compartments to open subcellular space, from protein-centered profiling to RNA/DNA-centered analysis, and from cell culture to animal. As this is not intended as a
The mechanism and evolution of enzyme-mediated proximity labeling. (a) Schematics of proximity labeling (PL) workflow. An enzyme (a biotin ligase or a peroxidase) is targeted to a specific subcellular location (e.g. mitochondrial matrix) via fusion with protein markers or signal peptides. Proximity is achieved through the in situ enzymatic synthesis of biotin-conjugated reactive intermediates, which subsequently diffuse away and react with nearby proteins/RNAs. The nanometer-scale action radius of the intermediates (shown as a red contour map) covers both proteins/RNAs that tightly associate with the bait and those that loosely interact in the same compartment, enabling PL to reach over multiple layers of protein–protein/RNA interactions. After cell lysis, biotinylated proteins are collected by affinity purification and characterized by mass spectrometry. Biotinylated RNAs are analyzed by high-throughput sequencing. (b) The mechanism and technology development timeline of PL. In the presence of H2O2, APEX (green, PDB 1V0H) converts biotin phenol to phenoxyl free radical, which reacts with the adjacent tyrosine residues. In the presence of ATP, BioID (cyan, PDB 2EWN) activates biotin into Bio-AMP, which reacts with lysine residues of neighboring proteins. The timeline describes a brief history of major APEX- and BirA-mediated PL techniques. Methods highlighted in green, blue, and pink refer to protein-centered, RNA-centered, and DNA-centered profiling, respectively.

comprehensive chronology of all PL applications, interested readers may find further information in several excellent reviews [12–15].

The evolution of promiscuous enzymatic labeling
The high spatial resolution of PL technique is achieved via both genetic targeting of the enzyme and the small action radius of the reactive intermediate. Depending on the nature of enzymes and chemical reactions, PL is broadly categorized as biotin ligase mediated or peroxidase mediated.

BioID (also known as BirA* [16]) is a 35 kDa *Escherichia coli* biotin ligase mutant R118G*BirA*. In the presence of ATP, BioID converts biotin into biotinyl-5'-AMP (Bio-AMP), which is subsequently released into the cellular
environment and reacts with the lysine side chain of nearby proteins [17](Figure 1b). Initially developed in 2012, BioID has been broadly applied to cultured mammalian cells [13], plant cells [18], mouse [19], yeast [20], etc., to profile proteomes in numerous subcellular structures, including the nuclear lamina [17], the nuclear pore complex [21], and centriolar satellites [22](Figure 2). In 2016, an improved smaller biotin ligase (BioID2) was developed from an *Aquifex aeolicus* enzyme to enable more-selective targeting of fusion proteins, leading to a better coverage of the nuclear pore complex components [23].

BioID/BioID2 requires several-hour labeling due to slow enzymatic kinetics [24]. Two strategies have been used to address this limitation. In the first one, Ting and coworkers applied yeast display-based directed evolution to enhance the catalytic efficiency of BioID. The resulting ligase mutant, termed TurboID, has up to 23-fold higher activity and reduces the labeling time window down to 10 min. Meanwhile, a truncation variant with reduced protein size and comparable labeling efficiency was introduced as miniTurbo [24]. In the second strategy, by analyzing the sequence and structural alignment of biotin ligases from different species, Khavari and coworkers introduced mutations to a *B. subtilis* biotin ligase to obtain a rationally designed PL enzyme, BASU [25]. When used to investigate the RNA-binding proteome (RaPID), BASU exhibited kinetics more than three orders of magnitude faster than BioID [25]. However, in another study, the activity of BASU was shown to be comparable with BioID/BioID2 in the cytoplasm [24].

Peroxidase-mediated PL offers even faster reaction kinetics by using free radical chemistry, as exemplified by horseradish peroxidase (HRP). Peroxidases could act on a wide variety of aromatic substrates, including luminol.

**Figure 2**

Representative applications of enzyme-mediated proximity labeling. (a) Primary cilia-targeted APEX2 identified several ciliary signaling molecules, including PKA, AMPK, and LKB1. (b) STIM1-fused APEX2 mapped the proteome of endoplasmic reticulum (ER)–plasma membrane junctions and uncovered an ER-localized transmembrane protein STIMATE. (c) Applications of APEX-mediated PL in mitochondria. (d) Interaction network and agonist response of GPCR were tracked by space- and time-resolved APEX PL (e–f) Applications of BioID to analyze the protein–protein interaction networks of two cytosolic membrane-less ribonucleoprotein complexes known as processing bodies (e) and stress granules (f). (g) A comprehensive proteomic profiling of centriolar satellites by BioID-suggested satellite assembly is independent of centrosome. (h) Profiling nuclear lamina proteome with BioID.
Emerging trends of enzyme-mediated proximity labeling. (a–b) From membrane-bound organelles to open-space compartments. Peroxidase-based PL was used to distinguish between the excitatory and inhibitory synaptic cleft proteome (a). Split enzymes were used to profile proteome at organelle contact sites (b). (c) From static to dynamic view of local proteome. Time-resolved APEX PL reveals dynamics of GPCR interaction network. (d) From protein-centric to peptide-centric analysis. Top: APEX PL with desthiobiotin-phenol whose reduced affinity to streptavidin (SA) facilitates recovery of labeled peptides. Bottom: APEX PL with alkyne-phenol allows functionalization of alkynylated peptides with azide-conjugated biotin bearing a UV-cleavable linker. Enriched peptides are released from streptavidin-coated beads via UV irradiation and identified by tandem mass spectrometry. Both approaches improve detection of labeling sites (e–f) From protein-centered to DNA-centered proteomic profiling. Proteins associated with specific genomic loci (e) or epigenetic marks (f) are labeled with dCas9-targeted APEX2 (e) or chromatin reader-fused BASU (f), respectively. (g) From protein-centered to RNA-centered proteomic profiling. The RNA-binding proteome is captured by BASU targeted via λN peptide/BoxB interaction (h–i) From proteomic to transcriptomic analysis. APEX PL is combined with protein-RNA crosslinking, via either formaldehyde or UV irradiation. Following purification of biotinylated protein–RNA complex, RNA could be identified by high-throughput sequencing (h). RNA could also be directly labeled by APEX PL (i).
for chemiluminescence [26], 3,3′-diaminobenzidine for electron microscopy contrast [27], and tyramine for protein labeling and signal amplification [28]. Because HRP is inactive when expressed in the cytosolic compartments [27], its applications have been restricted to the cell surface [11,29]. APEX is a 28 kDa engineered peroxidase that generates phenoxyl free radicals in living cells to label proximal proteins (Figure 1b). The first application of APEX in the mitochondrial matrix required only 1-min labeling and captured 495 proteins with >95% specificity, which helped reassign the proper localization of heme metabolic enzymes, PPOX and CPOX [30].

In 2015, through yeast display-based directed evolution, Ting and coworkers identified a point mutation of APEX that substantially improved its catalytic efficiency (APEX2), likely by reducing its susceptibility to H2O2-induced inhibition [31]. Interestingly, this mutation is conserved in secreted peroxidases including HRP; thus effectively turning APEX2 more HRP-like [31]. Over the past 8 years, APEX/APEX2 has been applied to a wide variety of subcellular compartments [15], including the mitochondrial intermembrane space (IMS) and outer membrane [32,33], the ER–plasma membrane contact [34], and primary cilia [35] (Figure 2).

**Toward higher spatial and temporal resolution**

Recently, the applications of PL has been extended from membrane-bound organelles to open subcellular space (Figure 3a). Initial efforts of PL in the IMS revealed high cytosolic background, particularly those proteins located near the mitochondrial outer membrane. To improve the spatial specificity of PL, a ratiometric quantitative proteomic workflow has been designed, where the enzyme is targeted to two neighboring subcellular compartments (e.g. IMS and cytosol) in two parallel PL experiments, respectively. The ratio of biotinylation extent (e.g. IMS/cytosol) is then used as a “cytological ruler” to measure the relative proximity of labeled proteins to either subcellular locations. Because the effects of protein abundance, local chemical environment, steric accessibility, etc., are cancelled out during the ratio calculation, this ratiometric PL approach has eliminated bias toward these factors, thus effectively creating a proximity contour map.

A prominent example of PL in open subcellular space is the study of synapse-specific proteome in cultured neurons, where biochemical fractionation is incapable of distinguishing excitatory synapses from inhibitory ones. By targeting HRP to the plasma membrane (not specific to synapses), excitatory and inhibitory synapses, respectively, ratiometric PL successfully identified 199 proteins at excitatory synaptic clefts and 42 at inhibitory ones. This example demonstrates that peroxidase-mediated PL is both applicable to primary culture and capable of dealing with small amount of material [11].

The high spatial resolution of PL has enabled the investigation of protein–protein interaction networks in LLPS-driven membrane-less organelles, such as SGs and processing bodies. In 2018, through fusing APEX2 to an SG marker G3BP1, Yeo and coworkers discovered a pre-existing SG protein interaction network in un-stressed cells and further identified stress-specific and cell type–specific SG subproteomes [36]. Coincidentally, Gingras and coworkers applied BioID to uncover 7424 unique proximity interactions and 144 protein components of cytosolic RNA granules [37].

To further improve the spatial specificity at organelle interface, a split enzyme strategy has been developed (Figure 3b). Enzymes such as BioID [38,39], TurboID [40], and APEX2 [41] are split into two catalytically inactive fragments which only recover biotinylation capabilities when they physically interact and reconstitute at membrane–membrane contact sites. These split versions of enzymes have enabled the specific profiling of proteomes at the ER-mitochondrial contact (split-TurboID) [40] and mitochondria-associated membrane (Contact-ID) [38].

The fast reaction kinetics of APEX has enabled mapping dynamic proteomic changes with high temporal resolution (Figure 3c). In 2017, APEX2 was fused to G-protein coupled receptors (GPCRs) to investigate signaling complex assembly upon ligand binding at minute-level temporal resolution. While Kruse and coworkers revealed the spatial and temporal clues to GPCR signaling and internalization kinetics stimulated by different ligands [42], Lobinger et al. captured the location and function of previously unknown GPCR network [43].

**Beyond protein-centered analysis in cell lines**

Target identification in PL traditionally focuses on protein-level quantitation (protein ID), where biotinylated proteins are first enriched and then proteolytically digested into peptides for tandem mass spectrometry (MS/MS) analysis [30]. As the majority of analyzed peptides are not tagged with biotin, this workflow is inefficient at characterizing the sites of biotin conjugation (site ID). Unlike protein ID, which is prone to false positives arising from nonspecific protein binding to affinity purification beads, site ID could unambiguously assign protein targets because the labeled peptides are directly detected by MS/MS. An additional benefit of site ID is to provide information of protein structural accessibility to enzymatic labeling, information that can be used to derive membrane
protein topology or protein complex conformation [44]. However, the tight interaction between biotinylated peptides and streptavidin beads has hindered site ID [45].

One solution is to weaken the interaction. Rhee and coworkers replaced the biotin moiety in the APEX substrate with desthiobiotin that has reduced binding affinity to streptavidin beads (Spot-ID) and applied this technique to map the topological direction of 135 mitochondrial membrane proteins [46] (Figure 3d). Similarly, the replacement of streptavidin with antibiotin antibody has resulted in more than 30-fold higher ID rate for biotinylated peptides, as demonstrated by Carr and coworkers [47]. Another solution is to substitute the biotin moiety in the APEX substrate with a clickable bioorthogonal functional handle, such as an alkyne group. Labeled proteins could be conjugated with an affinity tag containing a photocleavable linker, which facilitates the identification of labeling sites by MS/MS [48] (Figure 3d). When analyzing site ID data sets, one should be mindful of potential false negatives, as site ID reflects not only topology and complex geometry but also the presence and accessibility of particular target side-chains (i.e. lysine for BioID and tyrosine for APEX).

In addition to studying protein—protein interactions, PL has been applied to map proteins associated with specific genomic loci (chromatin-binding proteomes) (Figure 3e–f). Two methods (C-BERST [49] and GLoPro [50]) combine dCas9-based genome targeting and APEX2 labeling to discover proteomes bound to telomeres, centrosomes, and MYC promoters. Besides, ChromID has been introduced to obtain protein networks at DNA methylation and histone trimethylation residues, through fusing engineered chromatin readers to BASU [51]. The enzymes could be targeted to specific RNA molecules in a similar fashion to investigate RNA-binding proteome (Figure 3g). Fusions with bacteriophage MS2 (RNA-BioID) [52], λN peptide/boxB RNA (RaPID with BioID or BASU) [25], or CRISPR-Cas13 system [53] has allowed the identification of novel RNA-protein interactions in β-actin mRNA and Zika viral RNA.

More recently, two strategies have been used to extend APEX2 labeling to profiling the subcellular organization of RNA molecules (Figure 3h–i). In the first strategy, APEX2-mediated protein labeling was combined with protein-RNA crosslinking methods, either with formaldehyde (APEX-RIP) [54] or via UV illumination (Proximity-CLIP) [55]. While APEX-RIP performed well in profiling transcriptome of membrane-enclosed organelles, its spatial resolution is poorer in the cytoplasm. Alternatively, direct RNA labeling by APEX2 was achieved. APEX-seq developed by Ting and coworkers could profile subcellular transcriptomes at nine landmarks and correlate them to genome architecture and protein localization [56]. Ingolia and coworkers applied APEX-seq and APEX-mediated protein labeling to comprehensively understand the organization of translation-initiation complexes and SGs [57]. Through screening a panel of aromatic APEX substrates, the efficiency of APEX-seq was substantially enhanced by replacing biotin-phenol with biotin-arylamine probes [58].

Finally, proximity-dependent labeling has been applied not only in the cell culture but also in animals. BioID has been applied to discover inhibitory postsynaptic proteome in neonatal mice [19]. Tissue- and subcellular location-specific expression of APX in Caenorhabditis elegans has identified over 3000 proteins in the nucleus and the cytoplasm from four tissues [59]. To investigate the molecular mechanism underlying brain wiring, the cell surface proteome was profiled by HRP in the fly brain [60].

Outlook for proximity labeling

Given the rapidly expanding PL toolkits, one may wonder which method is better. The answer depends on the specific experimental setup, including temporal resolution, sensitivity to oxidative stress, probe delivery issues, etc. The subminute temporal resolution of APEX method remains unrivaled, and the requirement for \( \text{H}_2\text{O}_2 \) delivery provides a means for temporal gating, which is a prerequisite for studying dynamic processes such as ligand-triggered cellular signaling. However, if the temporal resolution is not a major concern, TurboID would be a good choice as it avoids the harsh treatment of \( \text{H}_2\text{O}_2 \). The dependence of BP probe delivery also poses a challenge to APEX application in live tissue [59,61]. In cases where both APEX and BioID are suitable, it would be informative to compare results from the two methods, which may complement each other as these enzymes target different amino acids. For example, in the mitochondrial matrix, 220 out of 495 proteins identified by APEX were also captured by TurboID [24,30], while in the ER membrane, 313 proteins were uncovered by both APEX (637 proteins) and TurboID (808 proteins) methods [24,33].

As PL techniques continue to evolve, future applications would likely benefit from integrating with other chemical biology techniques. For example, the combination of PL and photoactivation [62,63] could leverage the high temporal resolution of both methods, which would be particularly useful for studying subcellular proteomic changes when a specific cellular signaling transduction node is activated. We expect applications to studying the dynamic process of signaling complex assembly in the context of unfolded protein response, oxidative and heat stress, etc. in the coming years. Another possibility is to combine PL with protein/RNA
modification profiling. While the current analysis of PL focuses heavily on protein/RNA abundance at specific subcellular locations, future development may include quantitation of their chemical modifications and turnover rates (e.g. protein post-translation modifications, RNA epigenetic marks, etc.), which could further shed light on the function of organelles and protein assemblies in their native context.

Finally, the concept of proximity labeling may inspire the development of biocompatible photocatalysts. Like enzymatic labeling, photocatalysis is also characterized by multiple substrate turn-over and signal amplification. Earlier this year, MacMillan and coworkers used anti-body-conjugated iridium photocatalyst to generate carbene from a diazirine probe in the vicinity of specific receptor molecules. This method, called MicroMapping (μMap), was applied to map protein—protein interactions on the cell surface [64]. Similarly, our group has used genetically encoded photosensitizers and singlet oxygen-mediated nucleobase oxidation to label subcellular transcriptome in live cells with high spatial resolution [65]. Because the delivery of light is readily controllable and precise, both spatially and temporally, we anticipate more developments and applications of photocatalytic reactions in the future, as an important supplement to the PL toolbox.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements
We apologize to colleagues whose work was not cited because of the space limitations of this review. We thank Gang Wang from our lab for helpful discussions and acknowledge funding from the Ministry of Science and Technology (2016YFA0507600, 2017YFA0503600), the National Natural Science Foundation of China (91733131), the Natural Science Foundation of Beijing Municipality (5182011), the Interdisciplinary Medicine Seed Fund of Peking University (BMU2017MC006), and Li Ge-Zhao Ning Life Science Junior Research Fellowship.

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

* of special interest
** of outstanding interest


Engineered promiscuous biotin ligases with improved kinetics, TurboID and minTurbo, via yeast display-directed evolved direction. The high catalytic efficiency of TurboID reduces the labeling time window to 10 min and enables biotin ligase-mediated PL in flies and worms.


A new mutant of biotin ligase BASU was engineered from *B. subtilis* with faster kinetics and higher signal-to-noise ratio over *E. coli* BioID. The combination of 3N peptide/bioX RNA labeling system and BASU-mediated proximity labeling identifies proteins associated with specific RNA motifs, thus enabling the detection of RNA-protein interactions.


Development of APEX method for PL in live cells. Application of APEX in the human mitochondrial matrix identified 495 proteins with exceptionally high spatial specificity, leading to the reassignment of matrix localization of two metabolic enzymes in the heme biosynthesis pathway.


Introduction of APEX2, a catalytically more active peroxidase engineered by yeast display-based directed evolution. APEX2 resolved the localization of a calcium-binding protein MICU1 using electron microscopy.


Development of APEX-seq for the identification of local transcriptome via directly labeling RNA molecules with biotin phenol. Application of APEX-seq to nine subcellular compartments provides a comprehensive map of the spatial organization of different RNA types and isoforms.


Application of APEX-based PL to profile protein composition of stress granules under various stress conditions and in different cell types.

This study identified pre-existing protein–protein interaction networks in this highly dynamic, membrane-less and liquid-like structure.


BioID was used to systematically characterize 119 proteins related to various aspects of mRNA life cycle and functions. The pre-centric analysis uncovered the spatial relationship of RNA regulatory complexes and identified 144 proteins that constitute the core structure of stress granules and processing bodies.


Time-resolved APEX-mediated PL for studying the dynamic assembly of G-protein-coupled receptor-signaling complex at sub-minute temporal resolution; also see Ref. [43].


Time-resolved APEX-mediated PL for studying the dynamic assembly of G-protein-coupled receptor signaling complex at sub-minute temporal resolution; also see Ref. [42].


A newly designed APEX substrate, desthiobiotin phenol, enables the identification of topology direction of 135 mitochondrial membrane space proteins, due to a more efficient recovery of biotinylated peptides.


An anti-biotin antibody-based enrichment was introduced to dramatically improve the detection of biotinylation sites in enzyme-mediated PL.


C-BERST was developed based on CRISPR-dependent genome targeting and APEX-mediated proximity labeling to enable proteomic mapping at telomere and centromere loci, thus enabling the detection of DNA–protein interaction; also see Ref. [50].


GLoPro was developed based on CRISPR-dependent genome targeting and APEX-mediated proximity labeling to enable proteomic mapping at MYC promoter and telomere loci, thus enabling the detection of DNA–protein interaction; also see Ref. [49].


Introduction of APEX-RIP, the first application of APEX-mediated PL of RNA-protein interaction; also see Ref. [49].


Development of APEX-seq for the identification of local transcriptome via directly labeling RNA molecules with biotin phenol. Application of APEX-seq to nine subcellular compartments provides a comprehensive map of the spatial organization of different RNA types and isoforms.


The paper expanded the scope of APEX substrates. Seven novel aromatic substrates were synthesized and compared for their labeling performance toward protein, RNA, and DNA. Biotin-aniline was identified as the most efficient probe for RNA labeling and used for profiling the mitochondrial matrix transcriptome.


