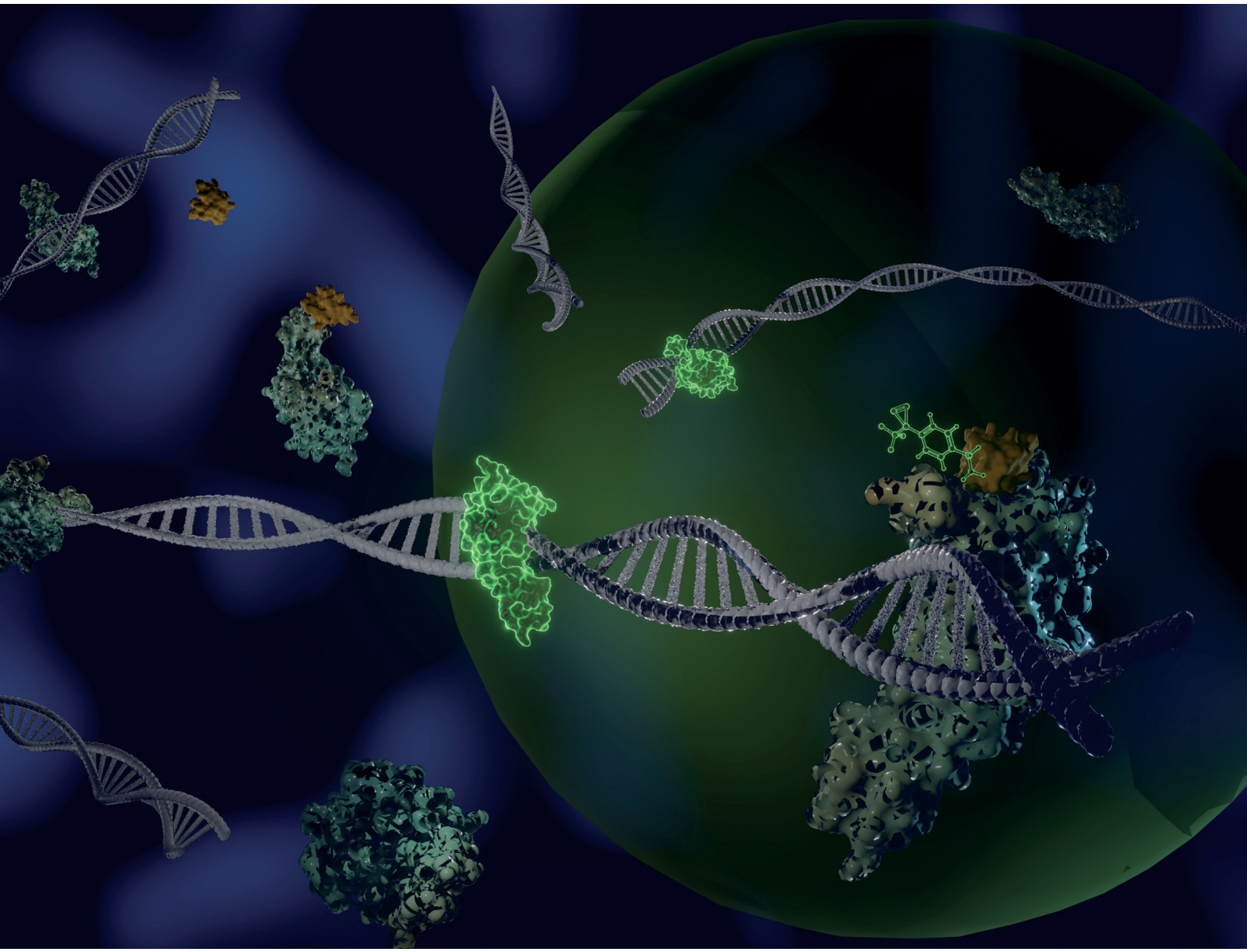
Volume 50
Number 5
7 March 2021
Pages 2895–3630

Chem Soc Rev

Chemical Society Reviews

rsc.li/chem-soc-rev



ISSN 0306-0012

ROYAL SOCIETY OF CHEMISTRY

TUTORIAL REVIEW
Tom W. Muir, David W. C. MacMillan *et al.*
Reactive intermediates for interactome mapping

Chem Soc Rev

TUTORIAL REVIEW

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2021, **50**, 2911

Reactive intermediates for interactome mapping

Ciaran P. Seath,[a] Aaron D. Trowbridge,[a] Tom W. Muir *[b] and David W. C. MacMillan*[a]

Received 1st November 2020

DOI: 10.1039/d0cs01366h

rsc.li/chem-soc-rev

The interactions of biomolecules underpin all cellular processes, and the understanding of their dynamic interplay can lead to significant advances in the treatment of disease through the identification of novel therapeutic strategies. Protein–protein interactions (PPIs) in particular play a vital role within this arena, providing the basis for the majority of cellular signalling pathways. Despite their great importance, the elucidation of weak or transient PPIs that cannot be identified by immunoprecipitation remains a significant challenge, particularly in a disease relevant cellular environment. Recent approaches towards this goal have utilized the *in situ* generation of high energy intermediates that cross-link with neighboring proteins, providing a snapshot of the biomolecular makeup of the local area or microenvironment, termed the interactome. In this tutorial review, we discuss these reactive intermediates, how they are generated, and the impact they have had on the discovery of new biology. Broadly, we believe this strategy has the potential to significantly accelerate our understanding of PPIs and how they affect cellular physiology.

Key learning points

1. Protein–protein interactions are traditionally difficult to identify.
2. Protein–protein and protein–nuclei acid interactions underpin many cellular processes.
3. Reactive intermediates such as activated esters, radicals, and carbenes can cross link with proteins.
4. Proximity labelling can be applied in many cellular contexts to investigate fundamental biological pathways.
5. Proximity labelling methods are based on chemical reactivity principles, such as solution half life.

1. Introduction

Proteins, along with RNA and DNA, comprise one of the three fundamental classes of biomolecules within organisms. The cellular proteome is understood to be vast, due to the potential for post-transcriptional and post-translational diversification, with each proteoform maintaining a unique structure and function.[1] This broad palette of biomolecules plays a role in almost every cellular process, providing a crucial link between gene and phenotype. This enormous diversity is further increased through interactions between a given set of proteins, called protein–protein interactions (PPIs), which make up its interactome.[2] Specifically, the spatial localization of these proteins is termed its microenvironment (Fig. 1).

Within discrete cellular microenvironments, PPIs play a crucial role in modulating cellular function and growth.[3] These interactions are varied and exist on different timescales, manifesting for example, as long-lived protein clusters, or as transient events promoted by external stimuli.[2,3] The dynamic nature of PPIs arises from the flexibility of peptide secondary structure and myriad post-translational modifications that can dramatically affect the binding surface of a given protein. Despite their great importance, methods to derive information about PPIs have been historically limited.[3] Traditional techniques to identify these interactors has centred around affinity purification-mass spectrometry (AP/MS), immunoprecipitation, and yeast two-hybrid methods.[4] AP/MS in particular has been enabling in recent years, wherein a tagged protein of interest (bearing a molecular affinity handle) is expressed in a cell and purified by affinity chromatography. In theory, interacting proteins will be co-purified and can be identified by mass spectrometry or western blotting. However, high-background affinity, low natural protein abundance, forcing lysis conditions, and low sensitivity of detection all remain formidable challenges. As such, transient or weak PPIs have, until recently, remained almost undetectable, leaving a vast gap in our knowledge of cellular biology. In light of these challenges, methods that are capable of labelling

[a] *Merck Center for Catalysis, Princeton University, Princeton, NJ, 08544, USA. E-mail: dmacmill@princeton.edu*
[b] *Department of Chemistry, Frick Chemistry Laboratory, Princeton University, Princeton, NJ, USA. E-mail: muir@princeton.edu*

biomolecules at a specific locus, termed proximity labelling, have emerged as powerful technologies for the identification of new PPIs, which have significant therapeutic implications in areas ranging from immune-oncology to neurodegenerative diseases.[3]

One strategy that has risen to prominence over the last two decades is the localized generation of high energy intermediates. These reactive entities have the ability to cross-link rapidly with adjacent biomolecules, generating a snapshot of the protein interactome, including transient PPIs. This approach manifests itself through the use of an antenna centred within the desired microenvironment of interest, which remains inert until an external stimulus, such as an exogenous small molecule or light, is applied. This "antenna" can be a protein, photoactivatable dye, or a single amino acid side chain.

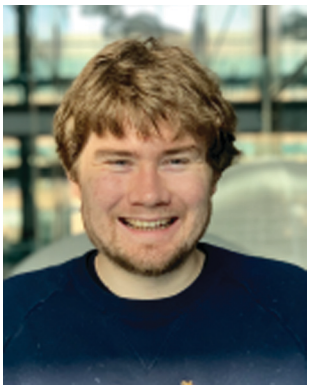
In this tutorial review we cover recent developments in this area including how these reactive intermediates are generated, their reactivity with biomolecules, and selected examples of how they have been exploited to aid our understanding of fundamental biology. Each labelling platform, by virtue of the reactive intermediate that they generate, is able to map comparatively larger or smaller areas of the desired microenvironment (10–1000 nm). Crucially, this is linked to the half-life of the intermediate within the cell, which in turn is governed by its fundamental chemical reactivity. Specifically, we will focus on carbenes, phenoxyl radicals, nitrenes, and activated esters, which have all seen broad use within microenvironment elucidation.

2. Basic principles of proximity labelling

The central tenet of proximity labelling is the localization of a small molecule or protein that can generate a reactive intermediate upon the application of an external stimulus. This species acts as an antenna, absorbing the external stimulus and delivering it to benign small molecule intermediates to form highly reactive

Ciaran P. Seath

Ciaran Seath completed his PhD in chemistry from the University of Strathclyde in 2017 where he worked under the supervision of Dr Allan Watson investigating chemoselective transition-metal catalysis. He then moved to Emory University as a postdoctoral researcher working in the laboratory of Professor Nathan Jui developing novel photoredox methodologies. Since 2019, Ciaran has worked with Professors David MacMillan and Tom Muir at Princeton University where he has been exploring new photocatalytic methods for proximity labelling.



Aaron D. Trowbridge

Aaron Trowbridge completed his PhD in chemistry in 2018 working under the supervision of Professor Matthew Gaunt at the University of Cambridge where he developed new photocatalytic methods for amine synthesis. He is currently a Marie Skłodowska–Curie in the laboratory of Professor David MacMillan at Princeton University, where his research focuses on the development of new proximity labelling technologies.

Tom W. Muir

Tom W. Muir is the Van Zandt Williams, Jr. Class of '65 Professor of Chemistry at Princeton University. He started his independent career at The Rockefeller University in 1996 and was subsequently appointed the Richard E. Salomon Family Professor and Director of the Pels Center of Chemistry, Biochemistry and Structural Biology before moving to Princeton in 2011. In 2020, he was elected to the American Academy of Arts and Sciences. His areas of research lie in the field of epigenetics, where he focuses on how changes to chromatin structure result in different cellular phenotypes.

David W. C. MacMillan

David W. C. MacMillan is the James S. McDonnell Distinguished University Chair of chemistry at Princeton University. He began his independent career UC Berkeley in 1998 before moving to the California Institute of Technology in 2000. In 2006, he became the A. Barton Hepburn Chair of Chemistry and Director of the Merck Center for Catalysis at Princeton University. In 2012, he was elected to both the American Academy of Arts and Sciences and the Royal Society. His research interests encompass a wide range of organic chemistry including the development of new areas in organocatalysis and photoredox catalysis.

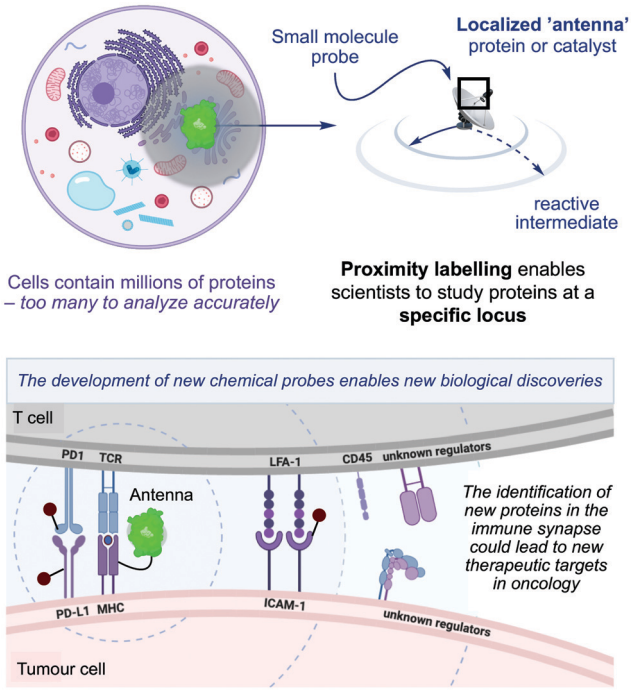


Fig. 1 Proximity labelling approaches can be used to identify protein–protein interactions within a defined cellular locus. These interactions can potentially regulate the function of the protein of interest. For example, proteins recruited to a synaptic junction could be viable therapeutic targets for immunology.

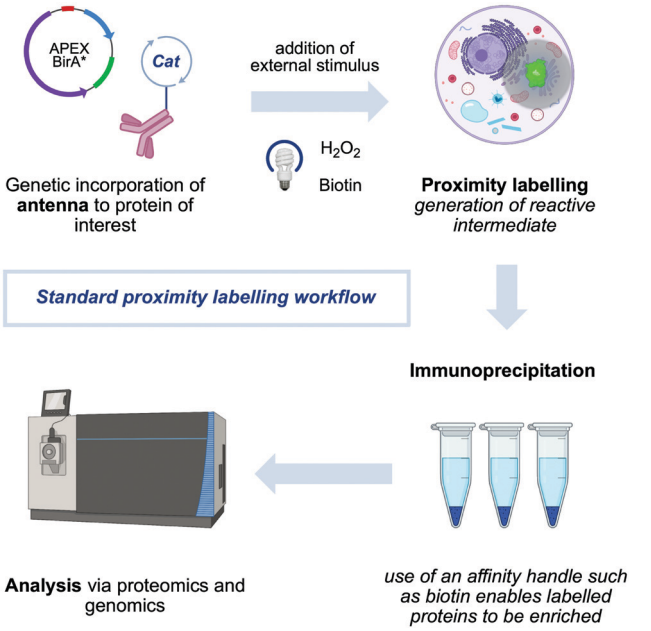


Fig. 2 Proximity labelling approaches follow a standardized workflow. The protein of interest is genetically engineered to express a protein fusion that can generate reactive intermediates. Labelled proteins or nucleic acids are enriched using streptavidin or immunoprecipitation. The enriched biomolecules are analysed by chemoproteomics or genomics.

species that can cross link with proteins only in the local environment. These intermediates typically contain an affinity handle, such as biotin, which facilitates selective enrichment of the labelled proteins. The strong interaction between the protein streptavidin and biotin allows users to extract all of the biotinylated proteins in the cell by treating the cell lysate with solid supported streptavidin. The extracted proteins can then be analysed by immunoblotting or by chemoproteomic analysis (Fig. 2).

The localization of such a species is typically performed by either genetic manipulation *via* protein fusions or through the action of an antibody that selectively localizes at a protein of interest.

One important aspect of proximity labelling is the effective labelling radius of the reactive intermediate, as this will define the precision and resolution of the data that is obtained. A less reactive probe will diffuse further and capture proteins over a greater distance than a more reactive probe, potentially giving rise to hundreds of potential 'hits'. Control over the distance scale is important as different applications will require dramatically different labelling radii.

3. Activated esters

Proximity-based biotinylation using the protein BirA* has become a cornerstone technology for interactome mapping. This method, termed BioID, employs mutant biotin ligase enzymes to catalyse the formation of electrophilic biotin AMP esters from the natural vitamin biotin. This reactivity can be localized to specific cellular locations using genetically engineered fusion proteins. When exposed to exogenous biotin, these cellular mutant biotin ligase enzymes release biotin-AMP, which reacts with lysine residues on surrounding proteins and can be detected through streptavidin enrichment and downstream proteomic analysis.

Since its introduction to the community, this methodology has been used in over 100 reports for the determination of cellular interactomes. In this section we will discuss the mechanism and development of BioID as a tool for interactome discovery, in addition to selected examples of its use to advance fundamental knowledge within cell biology.

3.1 Developments in BirA* mediated labelling

The first iteration of this method, BioID, employs a 33.5 kDa variant of the *Escherichia coli* biotin ligase BirA, which in its native form selectively delivers biotin to the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase. This proceeds *via* the combination of biotin and ATP to give biotinoyl-5′-AMP (BioAMP), which is held in the active site of BirA until it reacts with BCCP. In the active site, Lysine183 catalyses the reaction between biotin and ATP, facilitating the formation of BioAMP through a hydrogen bond between the phosphate and Arginine118.[5] In 2012, Roux and co-workers proposed that mutating these residues within the active site of BirA might erode the selectivity of the enzyme, resulting in premature release of BioAMP that could then react

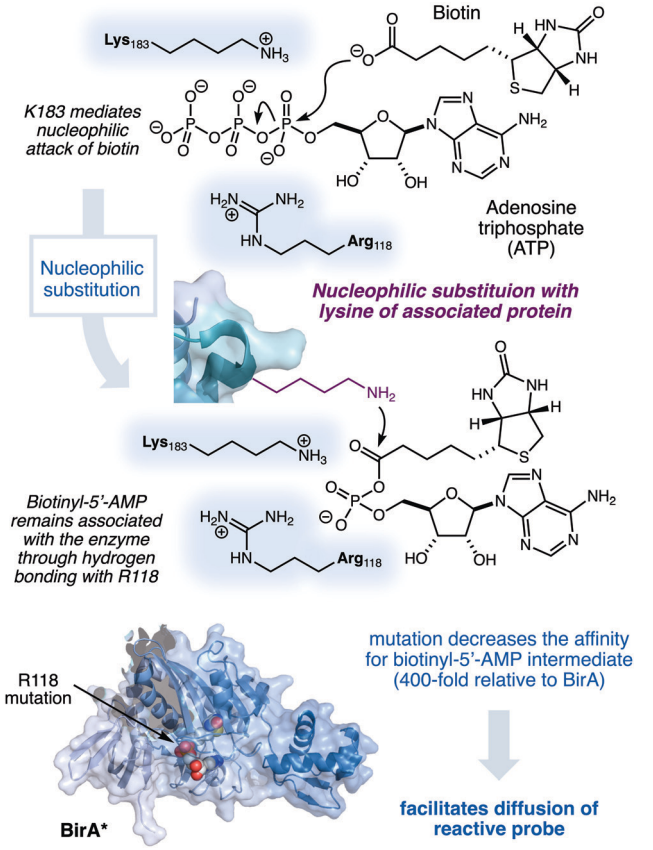


Fig. 3 BirA biotin ligase forms biotinyl-5′-AMP from biotin and ATP. Lysine 183 mediates the nucleophilic attack of biotin and arginine 118 locks the product in place. Mutagenesis of Arg118 leads to a promiscuous biotin ligase (BIrA*) that can be used to label lysine residues on proximal proteins. This technique is named BioID.

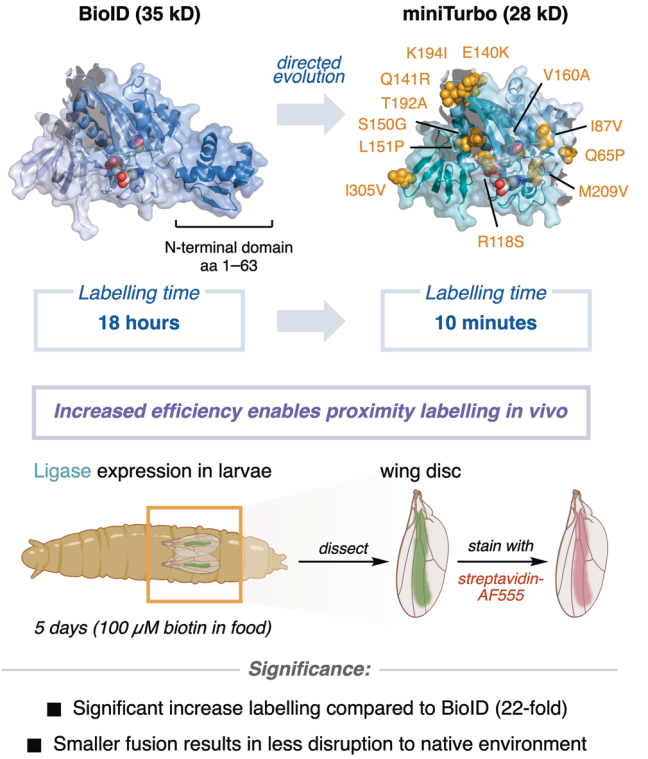


Fig. 4 The Ting group performed directed evolution on BirA* to improve labelling kinetics. The new biotin ligases, named TurboID and miniTurbo are amenable to *in vivo* labelling as demonstrated by biotinylation of larval wing disks in flies.

promiscuously with lysine residues in the surrounding environment.[6] Indeed, the variant R118G (named BirA*) lost two orders of magnitude binding affinity for BioAMP, allowing free diffusion of the activated ester out of the binding site. The proximity labelling reaction can be performed in living cells with BirA* fused to cellular proteins, where biotinylation of the local environment builds up over time. Indeed, the authors demonstrated time dependent biotinylation that could be determined by western blot analysis, with increasing levels of labelling over 72 h (Fig. 3).

Following this initial report, two mutated BirA variants (BioID2[7] and BASU[8]) were reported to provide modest improvements to the catalytic efficiency of BirA*; however, both new methods still required long labelling times which restricted their use for dynamic processes that require fine temporal control. In 2018, the Ting group addressed this limitation with the development of TurboID and miniTurbo.[9] Like BioID, these enzymes are mutated forms of the biotin ligase BirA. However, while BirA* only bears a single active site mutation, TurboID and miniTurbo contain 15 and 13 mutated sites respectively. These structural changes lead to a significant increase in activity, reducing labelling times from 18 hours to 10 min. Additionally, miniTurbo has 68 amino acids deleted from the N-terminus, rendering it smaller (28 kDa) than TurboID (35 kDa), without compromising catalytic activity. The authors identified these mutant biotin ligases through yeast-display based directed evolution after 29 rounds of selection. The increased labelling efficiency conferred by these two mutants has enabled proximity labelling *in vivo.* BioID, TurboID, and miniTurbo were selectively expressed in larval wing disks of flies (*Drosophila melanogaster*), which were fed a biotin rich diet for 5 days.[9] Dissection of the wing disks followed by streptavidin staining gave a quantitative measurement of biotinylation (Fig. 4). In this assay, both TurboID and miniTurbo exhibited a 22-fold increase in labelling when compared to BioID, demonstrating the power of BioID based approaches for *in vivo* labelling and the significant catalytic enhancements that can be imparted by directed evolution.

3.2 Selected applications of BirA* based proximity labelling

One of the benefits of BioID based approaches is their versatility with respect to the protein target. BirA* protein fusions can be constructed with almost any target of interest. One particularly significant example was reported by Penn and co-workers who used BioID to explore the interactome of the oncogenic transcription factor c-MYC.[10] This promiscuous transcription factor regulates the expression of thousands of genes, in particular those which control cell growth and proliferation. Its dysregulation in cancer is associated with poor prognosis and particularly

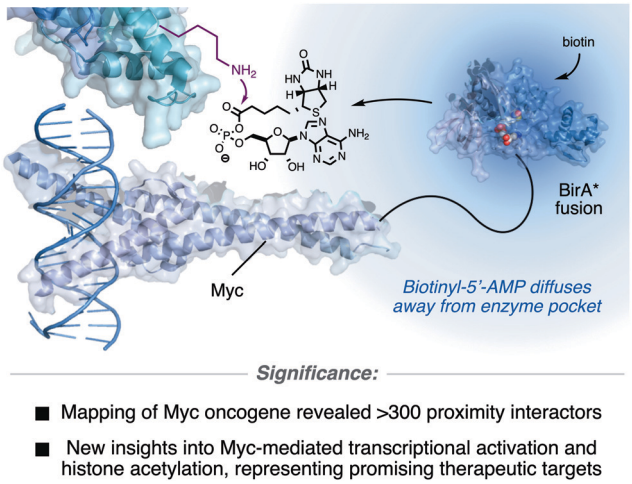


Fig. 5 A Myc-BirA* fusion protein was used to map proteins that interact with this oncogenic transcription factor. Performing BioID experiments with selected deletions of the myc protein reveals epitope specific interactions.

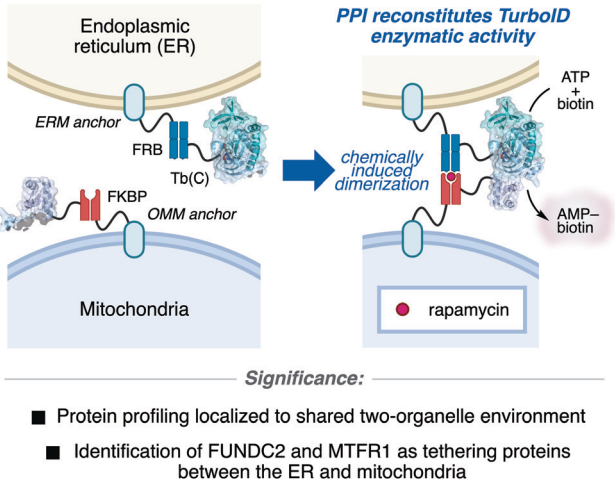


Fig. 6 Split TurboID was used by Ting and co-workers to profile PPis at the ER-mitochondrial membrane interface. Dimerization can be induced using a small molecule trigger, rapamycin, that forms a strong bond between FRB and FKBP.

aggressive tumour growth.[11] The action of this transcription factor is controlled by a complex web of protein–protein interactions, which have been shown to be modulated by mutations and PTMs to the structure of MYC. In this study, the authors investigated the interactomes of each of the six domains that make up MYC, termed Myc-boxes (MB), correlating PPIs to specific binding motifs. To achieve this goal, the group expressed MYC-BirA* proteins that contained specific sequence deletions and performed BioID experiments to compare their constituent interactomes with wild-type MYC-BirA*. These data demonstrated that many of the interacting proteins were dependent on one or more of the binding epitopes expressed on MYC, and their affinity to specific regions of MYC could be correlated to their primary function. Specifically, the region designated MBII was shown to be responsible for MYC dependent interactions with the epigenome. Deletion of this epitope led to a loss of a number of PPIs with histone acetyl transferases (HATs), which lay down epigenetic information that alters gene transcription (Fig. 5).

TurboID has also been used to examine the contact region between the endoplasmic reticulum and the mitochondria.[12] Standard approaches using BioID were not suitable to address this specific challenge, as localization of TurboID on either individual organelle would lead to significant non-specific labelling. In order to only profile proteins that are localized to the shared two-organelle microenvironment the Ting group developed a split TurboID strategy. In this method, TurboID is divided into two halves that are each fused to opposing sides of the two organelles. When the ER and mitochondrion come into close contact, the two halves TurboID recombine and can begin to label proteins in close proximity. To add an additional level of control, each half was fused to either the FKBP or FRB proteins, which form a strong heterodimer in the presence of the macrolide rapamycin. In this sense, proximity labelling is gated by both organelle proximity and a small molecule trigger. Under these conditions the authors found ~30 proteins that were enriched by proteomic analysis, and upon further validation were able to assign FUNDC2 and MTFR1 as having a tethering function between the ER and mitochondria (Fig. 6).[12]

Overall, BirA* based chemoproteomic discovery has changed the face of chemical biology, providing a tool for elucidating cellular interactions in the most complex of environments. The mild nature of this methodology (simple exposure to biotin) has facilitated extensive use *in vivo*,[13] and the improvements in the reaction kinetics granted by TurboID have significantly addressed limitations associated with its application on short timescales. However, this method exhibits low temporal control, which limits its use in more dynamic or stimulated systems. The activated ester species are also poorly reactive ($t_{1/2}$ ~ 5 min), which can lead to a wide labelling radius, particularly after saturation of neighbouring proteins with biotin. This issue can be somewhat attenuated through SplitTurboID, which offers rapamycin initiated biotinylation; however, this is not a general solution that can be applied in all cases.

4. Phenoxy radicals

The biocatalytic oxidation of phenols to phenoxy radicals through the action of peroxidase enzymes underpins some of the most widely used methods for cellular proximity labelling.[14] In these techniques, the peroxidase enzymes are situated at specific cellular locations where they catalytically generate phenoxy radical species. This enables cross linking to neighbouring proteins that can then be analysed to generate a local interactome. The crucial oxidation step is performed by metalloenzymes such as horseradish peroxidase or ascorbate peroxidase, which bear characteristic iron protoporphryin prosthetics (heme).[15] In this section, we will discuss the mechanism of activation, the physical characteristics of the reactive radical intermediates, and selected interesting examples of their use.

4.1 Generation of phenoxy radicals by horseradish peroxidase

Horseradish peroxidase catalyses the reaction of hydrogen peroxide and a phenol to water and a phenoxy radical, mediated by an iron porphyrin. This process begins with an Fe^{III} resting state, which is oxidized by hydrogen peroxide, *via* a peroxo-iron species, to give a molecule of water and oxo-ferryl (Fe^{IV}) radical cation. This species is then reduced to a neutral oxo-Fe^{IV} by phenolate to give an equivalent of phenoxyl radical, which is then free to diffuse into solution. The neutral oxo-Fe^{IV} can then oxidize another equivalent of phenolate to provide a molecule of H_2O and a second phenoxy radical (Fig. 7).

This general mechanism is operative in a number of peroxidases, tolerating a broad scope of phenol precursors. The phenoxyl radicals generated are long lived under aqueous conditions; the diffusion of these species is limited by radical dimerization, which occurs at a rate of 5×10^8 dm^3 mol^{-1} s^{-1} and equates to a solution half-life of ~1 ms.[16] These radicals cross-link with radicophilic biomolecules such as protein-associated tyrosine (*ortho* to the phenol) and nucleobases (at C-8) (Fig. 7). Additionally, the O-centred phenoxyl radical has a C-centred resonance structure, which leads to reactivity through either C2- or O, generating mixtures of products.

4.2 Selected applications of phenoxyl radical proximity labelling

Horseradish peroxidase (HRP) has been used within molecular biology for over 40 years for myriad applications including generating chemiluminescence for western blotting, biocatalysis, and ELISA assays. However, despite this broad usage, the first example of using HRP for proximity labelling was reported in 2008 by the Honke group.[17] The method, termed EMARS (enzyme-mediated activation of radical sources), employed HRP appended antibodies to direct labelling to specific cell surface proteins. Rather than the more typical phenol substrates, EMARS uses aryl azides, which form nitrenes following oxidation mediated by HRP. Unlike the radical-based reactivity observed through the oxidation phenols, electrophilic nitrenes are capable of undergoing insertion into electron rich C–H bonds (typical reactions with Trp). While the lifetimes of triplet aryl nitrenes in model systems are shown to be 0.1–1 ms, indicating a short labelling radius, substantially longer lifetimes on the order of minutes have been observed in more complex biological systems. This extended lifetime is thought to be caused *via* a ring-expansion of the aryl nitrene, through an intermediate azirine, to a 7-membered ketenimine (Fig. 8). This longer-lived reactive species undergoes reactions with amines to form azepines ($t_{1/2}$ ~ 5 ms to 1 s).[18] However, it remains largely unclear as to what effect ring substituents have on this pathway; although, polyfluorinated aryl azides are known to suppress azepine formation. Moreover, nucleophilic scavengers such as glutathione are capable of ablating labelling through reaction (and eventual reduction) with the azide moiety.

Nonetheless, the Honke group investigated molecular protein clusters around the transmembrane receptor EGFR, identifying an interaction with β-integrin by immunohistostaining.

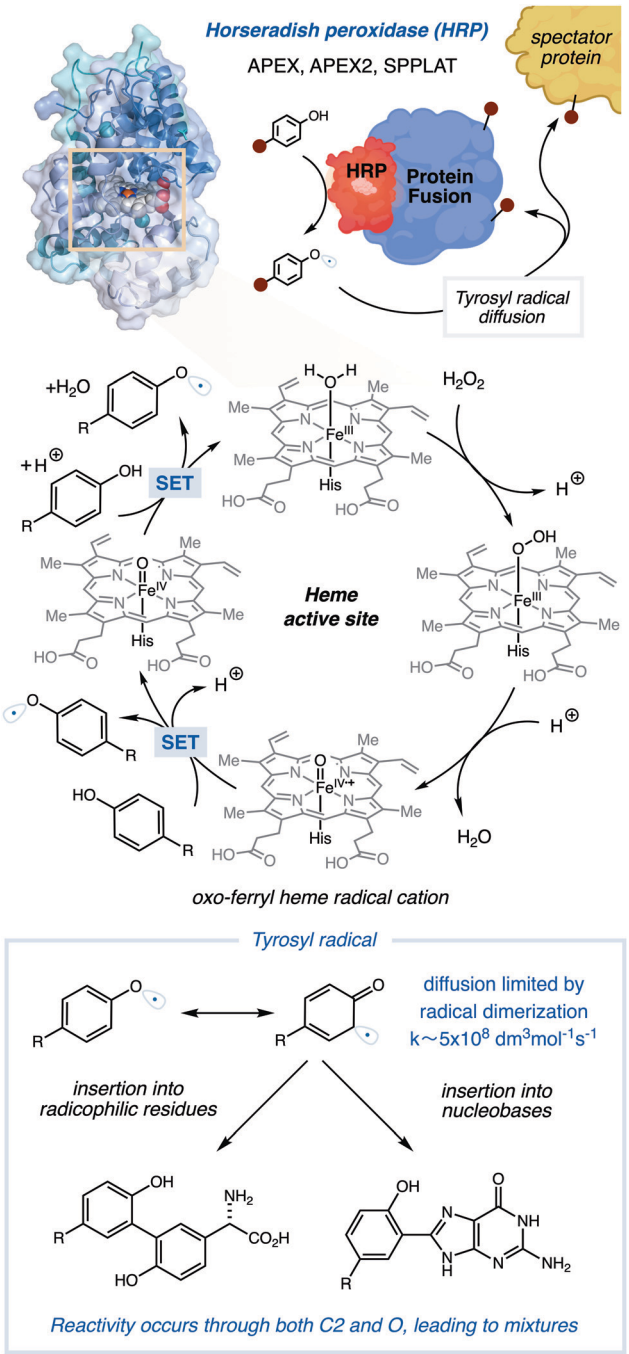


Fig. 7 Horseradish and ascorbate peroxidases can be used to activate phenols to form phenoxy radicals. These radicals can cross link with proteins and nucleic acids. This reactivity is enabled by a heme prosthetic within the active site of the peroxidase enzymes.

The group also investigated the labelling radius of this reaction using immunoelectron microscopy. In this case, the EMARS experiment was run on a nickel grid coated in mouse serum using an HRP-conjugated anti-mouse IgG antibody. The deposited labels were detected using streptavidin conjugated with 5 nm gold-colloids. Using this method, the authors estimated the labelling radius to be between 200 and

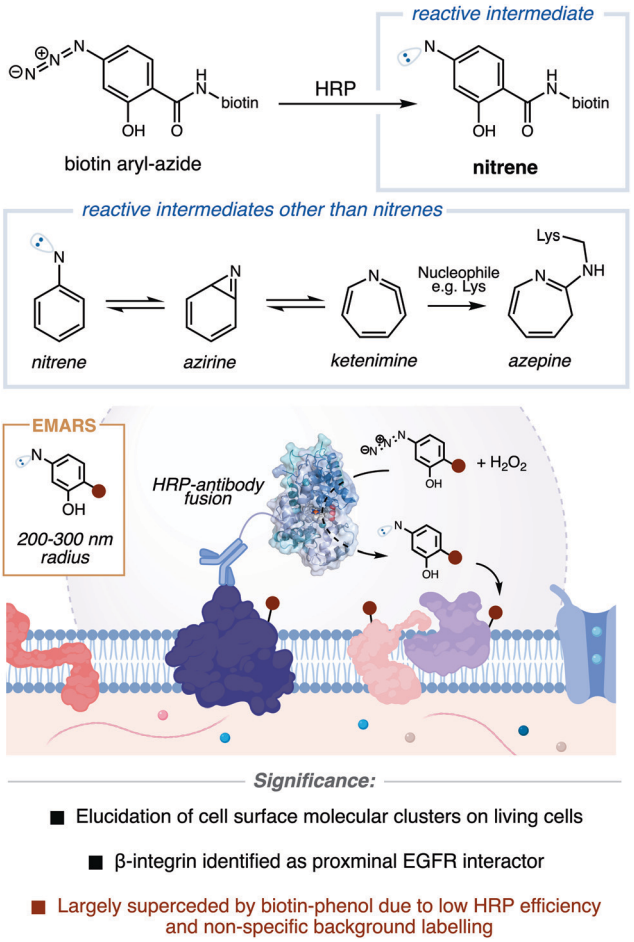


Fig. 8 EMARS utilises aryl azides for cell surface proximity labelling. EGFR directed labelling revealed β-integrin as a putative interactor.

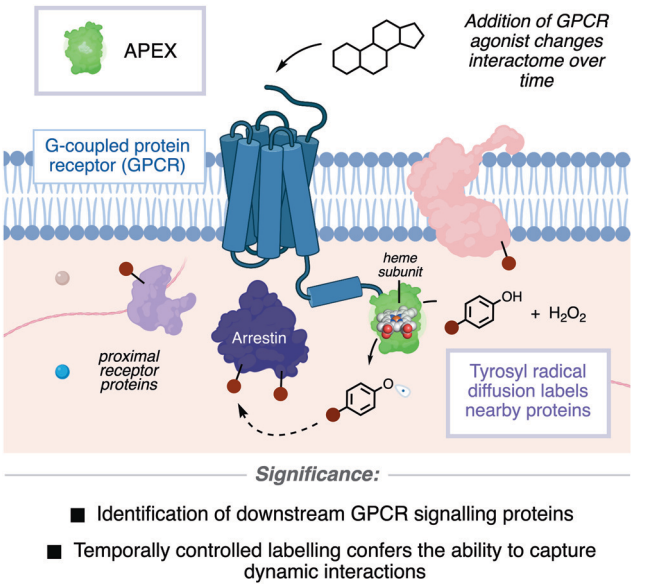


Fig. 9 The Kruse and Krogan groups used temporally controlled APEX2 proximity labeling to investigate ligand dependent GPCR signaling. APEX2 methodology provided 10s resolution and effectively tracked changes in signaling distribution between β-arrestin and GRKs.

300 nm, which approximately corresponds to the distances taken up by large assemblies of membrane proteins.

Since this initial disclosure, this platform has been expanded upon significantly by the groups of Jackson[19] (SPPLAAT) and Ting (APEX/APEX2).[20] SPLAAT is an antibody-based cell surface labelling method that built upon the EMARS technology. Rather than utilizing longer-lived nitrene intermediates, this technique employs phenoxyl radicals (generated from phenols) as the reactive intermediates, combining proximity biotinylation with LC/MS^2 analysis to generate interactome data.

While SPPLAT has been empowering, the most significant advances in this area have come from the development of APEX (enhanced ascorbate peroxidase) by the Ting group.[20,21] APEX is a soybean peroxidase that has been engineered to be more catalytically active, smaller, and able to be expressed as a protein fusion in human cell lines. In comparison to HRP, which is inactive within the reducing and calcium deficient nature of mammalian cells, APEX can be expressed in all cellular compartments, significantly broadening the potential for interrogating cell biology. Since the introduction of the method in 2012, it has been employed extensively to probe intracellular protein–protein interaction networks. The utility of this method was further improved in 2014 with the disclosure of APEX2, which has undergone further rounds of genetic engineering through yeast display evolution to exhibit faster reaction kinetics.[22]

Recently, the Kruse and Krogan groups independently described a particularly elegant examples in this field, deploying APEX2 to evaluate ligand dependent GPCR signalling.[23,24] In the example demonstrated by Kruse, the authors fused APEX2 to the C-terminal intracellular domain of angiotensin II type 1 receptor (AT1R) and evaluated alterations to the interactome in the presence of different ligands. The temporal control afforded by APEX2 was vital for success, enabling interactome discovery at both variable timepoints (10 s to 30 min) after exposure to a wide range of small molecule ligands (Fig. 9). This exquisite level of temporal control allowed the authors to track interactions over time between the AT1R and proteins that are known to initiate downstream signalling such as G-proteins, G-protein receptor kinases (GRKs) or ß-arrestin. This study in particular highlights the advantages provided by temporally controlled proximity labelling over BirA* and classical crosslinking methods, which struggle to capture dynamic interactions.

In addition to localization at fixed loci (*e.g.* membrane or organelle bound), APEX has recently been deployed in a dynamic manner using dCas9-APEX fusions, where APEX labelling can be performed at specific genetic loci (Fig. 10). In two simultaneous reports from Carr and Myers,[25] and Sontheimer,[26] APEX/2 could be directed towards a single gene by using an appropriate sgRNA molecule. For example, in the study from the Sontheimer group, dCas9 was directed to telomeric DNA in U2OS cells. Proteomic analysis following labelling showed that protein complexes that are known to regulate telomeres (including all 6 components of the shelterin complex) were highly enriched in

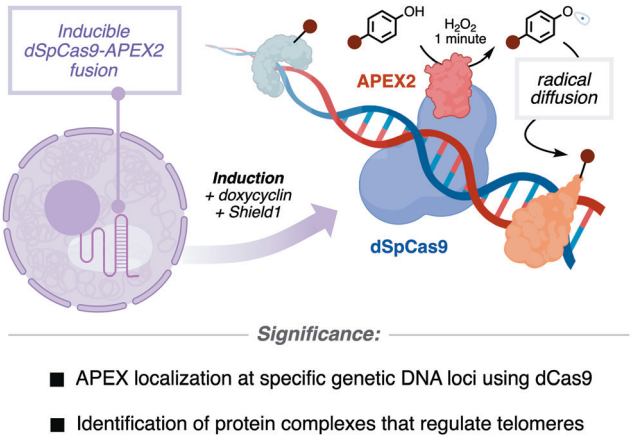


Fig. 10 dCas9-APEX conjugates can be directed to a specific genetic locus by a suitable sgRNA. Exposure to biotin-tyramide and H_2O_2 leads to localised generation of reactive phenoxyl radicals which can cross-link to nearby proteins. This technology allows the interrogation of the proteins involved in the regulation of a specific gene.

the analysis. Included in the two studies were targeting of four different genetic loci, and all showed orthogonal protein hits enriched in their analysis, demonstrating the level of accuracy that can be achieved using this method. The use of this methodology to unveil new interactions at physiologically important loci that remain underexplored is an exciting prospect.

A recent report from the Ting group expanded the method for the labelling of RNA to reveal subcellular transcriptomes at a broad variety of cellular locations.[27] In this method, APEX fusion proteins localized at specific organelles generate reactive phenoxyl radicals that cross link with neighbouring RNA. Affinity purification and sequencing provides a detailed cellular atlas of RNA sequence localization. The fine temporal control offered by the APEX technology enabled the RNA interactome to be probed under different physical stressors to provide support for a number of biological hypotheses. For instance, RNA is proposed to be transported to the mitochondria *via* microtubule assistance. To probe this, the authors performed APEX-seq at the outer mitochondrial membrane (OMM) under standard conditions and in the presence of microtubule polymerization inhibitor nocodazole (NOC). This experiment demonstrated that even after very short exposure times to NOC (3–30 min) the amount of RNAs localized at the OMM was significantly decreased, suggesting a strong correlation between RNA transport and microtubule polymerization (Fig. 11).

Shortly after this report, a study from the Zou group expanded the scope of substrates that are compatible with the APEX technology.[28] The authors proposed that the high oxidation potential and spacious active site of APEX2 could accommodate other probes that could provide differential reactivity (Fig. 12). An *in vitro* survey of these APEX2 probes with proteins, DNA, and RNA showed that biotin-phenol was indeed the optimal probe for labelling proteins. However, biotin-napthylamine was shown to be the most reactive probe for DNA labelling and biotin-aniline (bt-An) was optimal for RNA biotinylation. The probe was validated in HEK293T cells

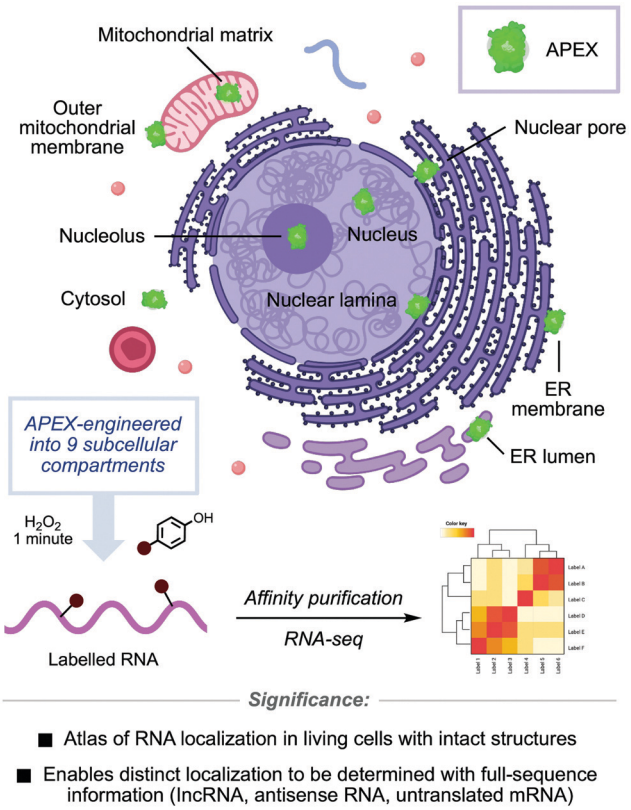


Fig. 11 Ting and co-workers demonstrated global RNA profiling in living cells using APEX2. Nine subcellular compartments were examined for RNA sequences. These data suggested roles for RNA in a number of biological processes such as protein localization.

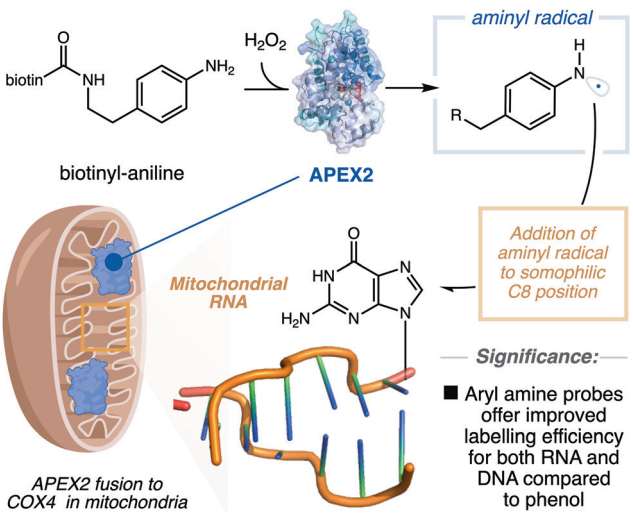


Fig. 12 A study from the Zou group showed that alteration to the probe structure can alter labeling efficiency between different biomolecules. Aniline as opposed to phenol was found to be optimal for labeling RNA. This aniline probe effectively labeled RNA in the mitochondria and nucleolus.

and in line with the work of Ting, was able to biotinylate RNAs in specific cellular compartments such as the mitochondria or

the nucleolus. This study demonstrates the value in structural modifications to the probe molecule, enforcing reactivity differences that can lead to altered function.

4.3 Ligand directed protein labelling

In addition to enzymatic generation with peroxidases, phenoxyl radicals have also been generated photochemically through the action of oxidizing photocatalysts. Building upon the pioneering work by Kodadek,[29] Nakamura devised a system whereby a Ruthenium photosensitizer was localized to the active site of a target protein (EGFR) *via* a tethered ligand (Gefitinib).[30] Upon visible light irradiation, the excited state photocatalyst can accept an electron from tyrosine residues proximal to the binding site to yield a tyrosyl radical. This reactive intermediate can be quenched by an excess of an electron-rich arene that is appended to biotin, forming a covalent bond that can be used to enrich the protein of interest (Fig. 13). In this sense, the protein target of the small molecule can be identified (named Target ID), which is a valuable tool for pharmaceutical development. This target ID methodology primarily provides data about the small molecule rather than the protein of interest and cannot be compared to the other labelling methods described in this review. Alternatively, the phenoxy radical intermediate can cross-link with a second tyrosine or other nucleophilic residue on an interacting protein, providing a stoichiometric cross-link that can be exploited for IP-MS detection. The blueprint laid down by Kodadek with this technology has led to numerous inventive methods that enable Target ID and PPI detection. This strategy offers an advantage when compared to proximity labelling approaches in that it captures only "true" PPIs (as opposed to proximal proteins); however, a number of restrictions are imposed by the fundamental design that may limit its utility in comparison to other labelling methods. For example, the method requires a highly selective ligand binding event and any PPI must occur proximal to both the ligand and a suitably reducing tyrosine residue. The field of ligand directed protein modification has advanced since Kodakek's seminal report, however, these examples lie outside the scope of this review, and have been discussed elsewhere.[31]

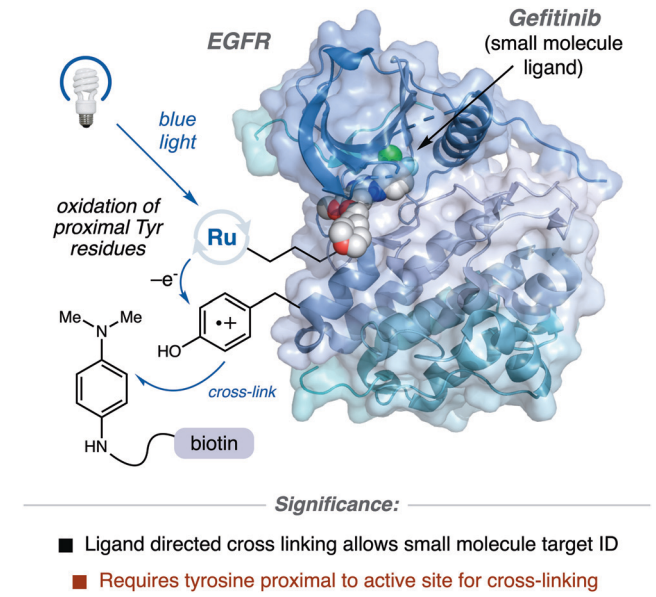


Fig. 13 Nakamura and others have reported ligand directed cross-linking using visible light photocatalysts tethered to small molecules. Upon excitation, the Ru photocatalyst is able to oxidize neighboring tyrosine residues that can cross-link with an excess of a competent radicophile that has a pendent biotin for enrichment (PDB: 4WKQ).

Broadly, phenoxyl radicals generated by engineered peroxidases have become a cornerstone method for detecting PPIs. The rapid reaction kinetics (~1 min exposure) allow for fine temporal control, and the radical species are able to cross-link to all types of biomolecules. In combination, these attributes have been extremely enabling. However, limitations in this method remain: (a) the generation of the phenoxyl radicals with H_2O_2 is relatively harsh, which may disrupt cellular processes that respond to reactive oxygen species (such as DNA damage and repair mechanisms), and (b) this strategy is not yet broadly applicable to *in vivo* applications (in comparison to BirA* based methods).

5. Diazirines

Diazirines are three membered rings, comprised of two nitrogen atoms bonded to the same carbon atom. These strained heterocycles are one of the most frequently employed probes within chemical biology. Upon exposure to UV light, these functional groups lose a molecule of nitrogen to generate a highly reactive carbene. These carbenes are the basis for the broad utility of diazirines within chemical biology, where they are able to insert into neighbouring C–H or X–H bonds, forming a stoichiometric cross-link. This strategy was initially employed in 1973 for photoaffinity labelling (PAL) by Knowles who identified diazirines as an ideal reagent for PAL, as this class of molecules exhibited improved stability with respect to diazo species and better reactivity than nitrenes.[32] Following this initial report, diazirines have been employed extensively as photocrosslinkers over the last 50 years. Here we will discuss mechanistic aspects of diazirine activation, methods for their incorporation, and recent innovative examples of their use in chemical biology.

5.1 Generation of carbenes through photosensitization

The photosensitization of diazirines proceeds through their exposure to 350 nm light, promoting the diazirine from a ground state singlet to its singlet excited state. This occurs through a $\sigma \rightarrow \pi^*$ transition where electron density from the σ_{C-N} bond is transferred to the $\pi^*_{N=N}$ bond.[33] Subsequent homolytic cleavage of the remaining σ_{C-N} bond gives molecular nitrogen and a singlet carbene. Intersystem crossing of the ground state singlet carbene to its triplet state (T_0) may occur in cases where the triplet state is of lower energy than the corresponding singlet.

In terms of reactivity, the singlet carbene displays polar reactivity, readily inserting into C–H and X–H bonds, found in all amino acids, in a concerted manner. This residue agnostic behaviour is highly desirable for proximity labelling

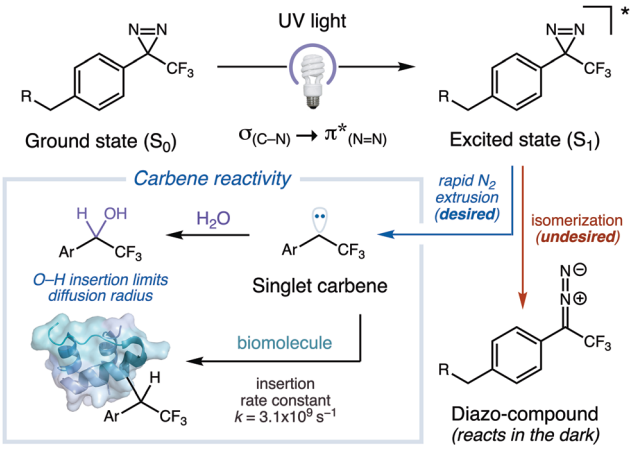


Fig. 14 Diazirines can be activated upon exposure to UV light to form carbenes. UV excitation promotes the singlet ground state into the singlet excited state which following extrusion of nitrogen provides a singlet carbene. Singlet carbenes are highly reactive and will cross link with biomolecules and also insert into water molecules, limiting diffusion. Undesired isomerization of the singlet excited state leads to diazo species which reacts more slowly, leading to background reactivity.

applications, making the singlet carbene the species of choice for practitioners. However, this reactivity also enables insertion into molecules of water leading to hydrated products, and consequently leading to very short solution half-lives ($t_{1/2}$ < 1 ns).[34] Conversely, triplet carbenes display radical reactivity, requiring a two-step mechanism to cross-link. This either proceeds through addition to a suitable π-system, or *via* abstraction of a weak X–H bond followed by radical recombination (Fig. 14). These mechanisms proceed at lower rates than the corresponding singlet carbenes and preclude reactivity with water, leading to a longer solution half-lives ($t_{1/2}$ = 2 μs).[34] Triplet carbenes are not considered to be residue agnostic as they are restricted to radicophilic residues such as tyrosine, histidine, cysteine, and tryptophan.

5.2 Chemical bait and trap methodologies

Chemical bait and trap, broadly defined, is a method for investigating the interactome of a specific epitope, or "bait" present on a biomolecule. The bait induces interactions with other biomolecules that can be captured by a neighbouring photoaffinity probe, termed the "trap". Proteomic analysis of the resulting crosslinked mixture reveals the microenvironment around the epitope of interest. Successful implementation of this strategy requires site specific incorporation of a photoaffinity tag (such as a diazirine) into a position that is proximal to the bait in order to capture all relevant interactions. While this is facile to carry out with recombinant protein,[35] which has been reviewed elsewhere, few approaches have been reported in a cellular context. Here, we will focus on examples of diazirine incorporation for the purposes of sampling the protein microenvironment.

5.3 Amber suppression

Incorporation of photoreactive amino acids into proteins in live cells was first demonstrated by replacing a canonical amino acid in the growth medium with a diazirine bearing variant, *i.e.* replacing methionine for PhotoMet or leucine for PhotoLeu.[36] This method relies on the cell's natural metabolism to recognise the amino acid and incorporate it into proteins, leading to uniform, global incorporation into the proteome. While effective in some cases, it does not allow positional selectivity, which is often required to pick up interactions at a specific epitope or domain. Additionally, the effect of global incorporation of an unnatural amino acid on cellular physiology is unknown and could lead to false positive or negative interactions. Following this, Schultz and co-workers were the first to report the site selective incorporation of a diazirine-bearing unnatural amino acid in live cells.[37] The authors employed their amber suppression technology to incorporate 4′-[3-(trifluoromethyl)23*H*-diazirin-3-yl]-L-phenylalanine (TfmdPhe) in *Escherichia coli* (*E. coli*) cells. The labelling radius of this diazirine probe was estimated to be approximately 6 Å, based on the distance from the base of the amino acid to the site of cross-linking. This short-range labelling technology has been expanded significantly in recent years, incorporating a range of diazirine bearing amino acids that exhibit different labelling radii (up to ~14 Å) to investigate a number of biological questions. This general strategy facilitates the cross linking of PPIs that occur on short distance scales and can be broadly adapted throughout the proteome. A recent example of this method was disclosed by Kleiner *et al.* who employed the Schultz method to incorporate the photoactivatable amino acid 3′-azibutyl-*N*-carbamoyl-lysine (AbK) onto the N-terminal tails of histones H3 and H4 in order to capture chromatin dependent PPIs at various stages of the cell cycle (Fig. 15).[38] This method unveiled a number of new interactions that were dependent on cellular context, such as cell cycle dependence (interphase *vs.* mitosis), or whether the protein is soluble or part of a larger histone complex. For example, the authors detected a weak interaction (50–100 μM) between H3 and UBR7 that is only present with soluble H3 (not observed in chromatin bound H3). This interaction was postulated to be involved in a ubiquitin dependent degradation of soluble histone H3.

5.4 Cleavable diazirine cross-linkers

One major challenge that remains in this field is the deconvolution of protein cross links by MS analysis, particularly when trying to uncover structural information of a new PPI. To counter this, cleavable diazirine linkers have been designed that allow the cross-link to be severed either before or during MS analysis. One such example from the Chen group, described the use of a selenium-based diazirine (DiZHSeC) that can be site specifically incorporated into native proteins *via* the amber suppression strategy.[39] After photo-crosslinking, the carbon–selenium bond can be cleaved to release the "prey" proteins from the (more abundant) "bait". This strategy removes the bait from the proteomic analysis, reducing non-specific background derived from an excess of the bait (Fig. 16). Additionally, cleavage of the C–Se bond allows single site determination of the site of cross-linking by removing the bait peptide from the analysis. A second strategy is to use MS cleavable cross-linkers that display a distinctive fragmentation pattern by MS/MS.[40] This strategy has enabled proteome wide analysis of crosslinking patterns at the

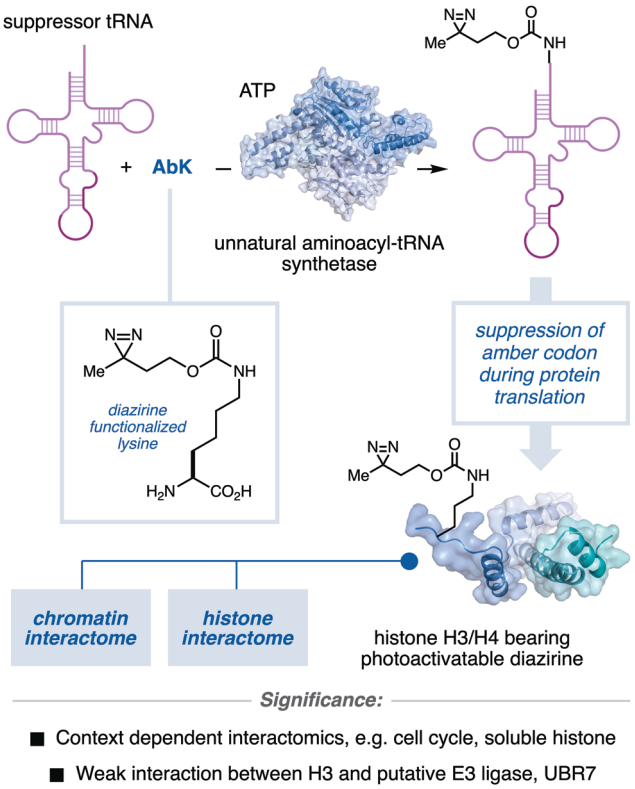


Fig. 15 Kleiner *et al.* used the Schultz amber suppression method to incorporate diazirine bearing amino acids into histone tails. Exposure to UV light at various stages of the cell cycle provided carbene intermediates that cross-linked to interacting proteins, revealing context dependent interactomics.

amino acid level by reducing the computational strain imposed by the *n*-square problem.[41]

5.5 Split intein

Protein semi-synthesis has recently emerged as a viable and powerful alternative for the elucidation of cellular microenvironments. In this approach, one half of an ultrafast split intein is fused to the protein of interest and its partner, which bears the delivery cargo, is prepared using chemical synthesis.[42,43] Recombination and protein trans-splicing proceed in cell to provide a new semi-synthetic protein. Solid phase peptide synthesis of the intein partner allows facile incorporation of photoactivatable unnatural amino acids, post-translationally modified residues, and affinity handles such as biotin. The flexibility of this approach allows for the ready incorporation of different "bait". This allows comparisons of the interactomes of native proteins with their mutated counterparts. In 2020, Muir and co-workers disclosed the first example of this method *in situ* (as opposed to *in vitro* with recombinant protein) in the context of chromatin interactomics (Fig. 17).[44] The authors interrogated the interactions associated with histone post-translational modifications (hPTMs) on H3 in HEK293T cell nuclei. By synthesizing an intein that contained a diazirine proximal to the hPTM of interest the interactome of a single amino acid can be revealed. A biotin affinity tag at the terminus of the peptide was used for

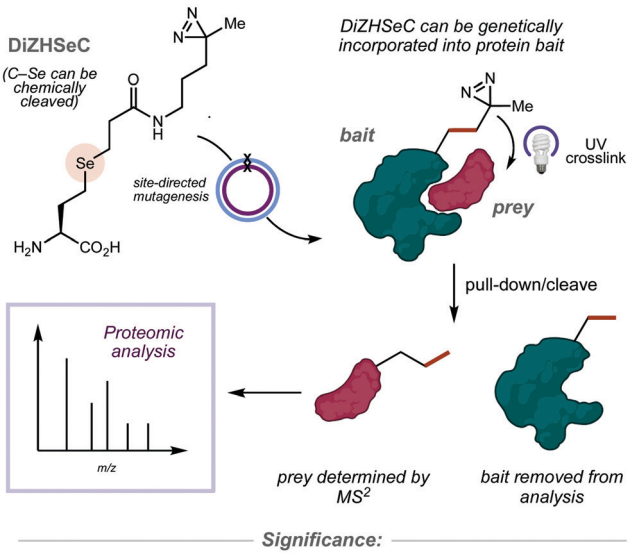


Fig. 16 Cleavable cross-linkers can enable high resolution chemoproteomic analysis by eliminating background from non-specific binding to the bait protein. This technique has been used to examine the cross-linking site at the amino acid level, providing structural information about protein protein interactions.

downstream pull-down proteomics. One particular example from this study compared the interactomes of hPTMs of H3K4 (H3K4wt *vs.* H3K4me1 or H3K4me3), identifying many of the known binding partners of these histone PTMs in addition to other hitherto unknown protein interactors. In comparison to the amber suppression method employed by Kapoor,[38] which probed how interactions change based on greater cellular changes (*e.g.* mitosis), this intein technology enabled the

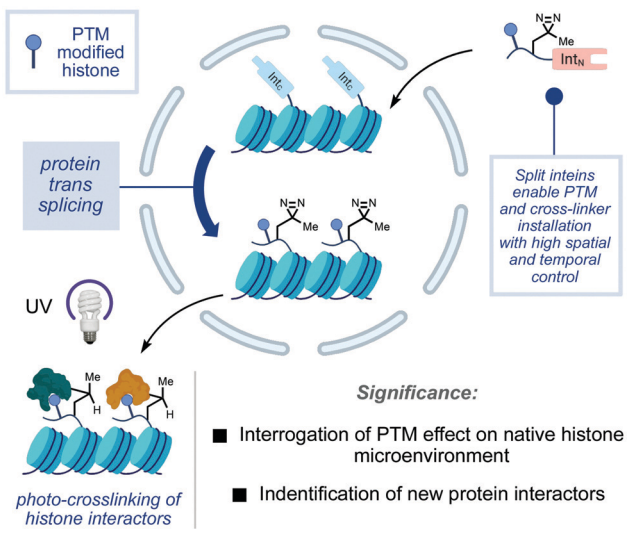


Fig. 17 Burton *et al.* demonstrate PTM dependent interactome mapping using intein trans-splicing. Semi-synthesis of histones in nuclei with diazirines located on the H3.3 peptide tail enables photocrosslinking upon UV activation in a native context.

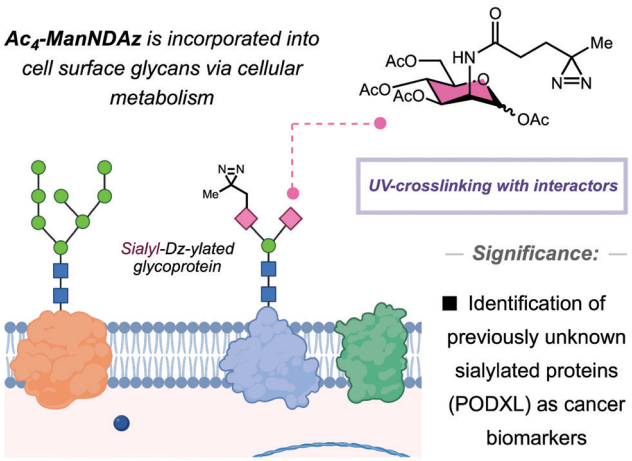


Fig. 18 A diazirine containing mannose derivative is incorporated into cell surface glycans by the cellular sialic acid biosynthetic pathway. These glycans can then cross-link with interacting proteins upon UV irradiation.

elucidation of how the histone microenvironment is affected by subtle chemical changes in peptide structure such as the presence of hPTMs.

5.6 Protein–glycan interactions *via* modified sialic acids

Sialic acid-based interactions play a number of roles on the cell surface and have been implicated in cancer immune evasion. In order to capture the interactions that underpin these biological roles, the Kohler group, building upon early work by Bertozzi,[45] elegantly demonstrated the incorporation of diazirines into cell surface glycans. This was achieved through the synthesis of ManNDaz, a mannose derivative that bears a pendent diazirine (Fig. 18) and can be metabolically incorporated using the cell's sialic acid biosynthetic pathway to provide diazirine functionalized cell surface glycans. Upon exposure to UV light, the diazirine handle decomposes to a carbene which is able to cross-link with interacting proteins. Using this method, the group was able to identify CD22, a known interactor,[46] and in a related system, identify previously unknown sialic acid interacting proteins, such as podocalyxin (PODXL).[47] Expression of PODXL is known to correlate with cancer aggressiveness and can be used as a predictor of breast cancer progression.

5.7 Incorporation of photocleavable linker

While empowering in many contexts, chemical bait and trap methods are less well suited to mapping larger microenvironments that contain multi-protein complexes or weak long-range interactions. The inherent labelling radius imposed by the length of the linker employed limits the capture of proximal proteins that are not situated within range of the epitope under examination. A recent development from the Moellering group has sought to extend this radius.[48] Their strategy is based on a dually activated probe molecule, which contains a diazirine and a photocleavable linker. Both components are activated by UV irradiation; however, if the photocleavable linker is preferentially excited (and subsequently cleaved) then it allows the diazirine molecule to diffuse through solution before being activated itself. In theory, such a platform could be tuned to provide variable labelling radii, all based upon the relative half-lives of each separate component. For example, a rapidly degraded photolabile group would provide a wider radius than one which is cleaved at a similar rate to diazirine activation.

The authors developed their probe using a diazirine that exhibited a half-life ($t_{1/2}$ = 55 s) in combination with a nitroveratryl photolabile group which displayed a shorter half-life ($t_{1/2}$ = 45 s). The probe molecule is localized in the cell using a SNAPtag fusion. The SNAPtag protein can be appended to any protein of interest and forms a covalent bond with a benzyl guanine motif that is present on the small molecule probe. The final part of the probe design is a biotin handle that can be used for streptavidin enrichment (Fig. 19). They exemplified their method by examining the microenvironment of KEAP1, which is a sensor protein that plays a role in cellular antioxidant response signalling. The SNAP-KEAP1 fusion protein was transfected into HEK293T cells before being treated with the trifunctional probe (Fig. 19). Irradiation with 350 nm light, followed by cell lysis, streptavidin bead enrichment, and proteomic analysis yielded the putative interactome. The fine temporal control enabled by this light driven technology was then harnessed to study how the KEAP1 microenvironment is altered in response to external stimuli. The authors demonstrated that the KEAP1

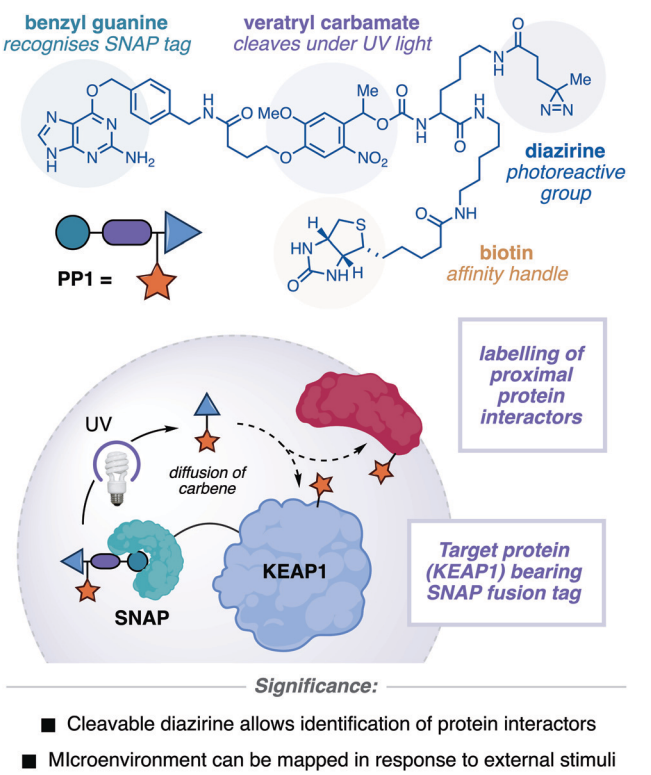


Fig. 19 A photocleavable linker strategy was employed by Moellering and co-workers to achieve light driven proximity labelling of the KEAP1 interactome under conditions of oxidative stress. The multifunctional probe is localized *via* a SNAPtag fusion protein to KEAP1. UV irradiation first leads to cleavage of the nitroveratryl group before diazirine activation, leaving a short diffusion half-life that allows labelling of proximal proteins.

interactome was significantly altered in response to oxidative stress (*t*BuOOH). Significantly, this represents an experiment that could not be achieved *via* any other method; BirA*-based labelling techniques to do not provide the appropriate level of temporal control, and APEX2 cannot accurately assess the effect of oxidative stress as it itself is activated by peroxide.

5.8 Catalytic activation of diazirines

One of the primary challenges with all known diazirine-based methods has been their rapid reactivity with water, leading to hydrated products. While this is vital to restrict diffusion and maintain a tight labelling radius, it also leads to very low labelling efficiencies rendering analysis of low abundance proteins extremely challenging.

Recently, the MacMillan group in collaboration with Merck disclosed a new methodology for the elucidation of cell surface microenvironments that aimed to address this fundamental limitation.[49] Their technology, termed μMap, exploits the ability of iridium photocatalysts to selectively absorb visible light and deliver it in the form of chemical energy to its immediate surroundings – generating transient high-energy intermediates. In this case, photoexcited catalysts are able to sensitize diazirine molecules, which subsequently release nitrogen, forming carbenes only in the vicinity of the iridium catalyst. This platform is based upon a fundamental photochemical process called Dexter energy transfer.[50] In this mechanistic event, a suitable photosensitizer in its triplet excited state can transfer energy to a ground state molecule, returning the photosensitizer to the ground state, and exciting the small molecule to its corresponding triplet state (Fig. 20). For this to occur, the photosensitizer must have a higher triplet energy than the small molecule substrate. Aryl diazirines have been shown to have a triplet energy of ~60 kcal mol^{-1},[50] which corresponds to the upper limit of conventional iridium-based photoredox catalysts. This energy transfer step only occurs within 1 nm, which when combined with the short solution half-life ($t_{1/2}$ < 2 ns) exhibited by carbenes, enforces a tight labelling radius (approx. 4–5 nm). Overall this activation mode delivers a proximity labelling methodology similar to HRP and BirA*, but with a more precise radius of labelling that can be used to profile distinct microenvironments on the nm scale (as opposed to μm). In this report, antibodies decorated with iridium photocatalysts were used to localize this energy transfer mechanism to specific cell surface proteins. Subsequent irradiation of the cells with blue light in the presence of excess diazirine resulted in the continuous deposition of tags within a very narrow radius of the antibody, leading to high-resolution mapping of the protein microenvironment. Analogous to previous reports (BioID, APEX2), the diazirine tags contained biotin affinity handles for facile downstream processing using chemoproteomics. The team exemplified this technology by exploring the microenvironment around PD-L1, a crucial cell-surface marker upregulated in many cancers (Fig. 21). In a single experiment, the team were able to identify a number of known interactors and suggest putative novel interactions that could not be identified through lower resolution methods such as immunocoprecipitation. Significantly, the authors compared their labelling method to existing HRP based methods (EMARS, SPPLAT) and found a significant increase in the definition of the proteomics data. One demonstration of this new methodology was in the context of the immune synapse, where authors showed that they could label both sides of a B/T cell synapse at the PD-1/PD-L1 immune checkpoint. In comparison to peroxidase-based methods, which were shown to label both cells non-specifically, confocal microscopy indicated that μMap only labelled the PD-1 bearing cell at the synaptic junction. This experiment provides proof of concept

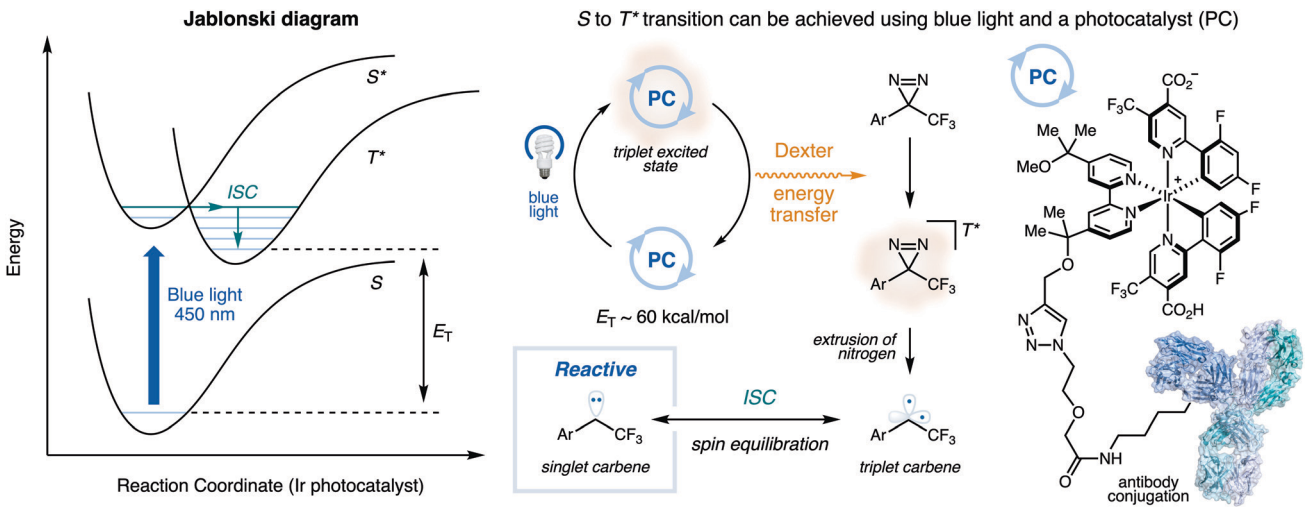


Fig. 20 Left: A Jablonski diagram describing the alternate pathway for diazirine activation described by MacMillan (blue arrow) that is driven by triplet excitation using blue light. Right: Irradiation of an Ir-photocatalyst with blue light leads to the triplet excited state which can undergo Dexter energy transfer with a diazirine to provide the diazirine triplet excited state which releases nitrogen to give a triplet carbene. Spin equilibration of this carbene to the singlet leads to residue agnostic cross-linking. The Ir-catalyst can be localised to specific cell membrane proteins using antibodies, to catalytically label proteins in the surrounding microenvironment.

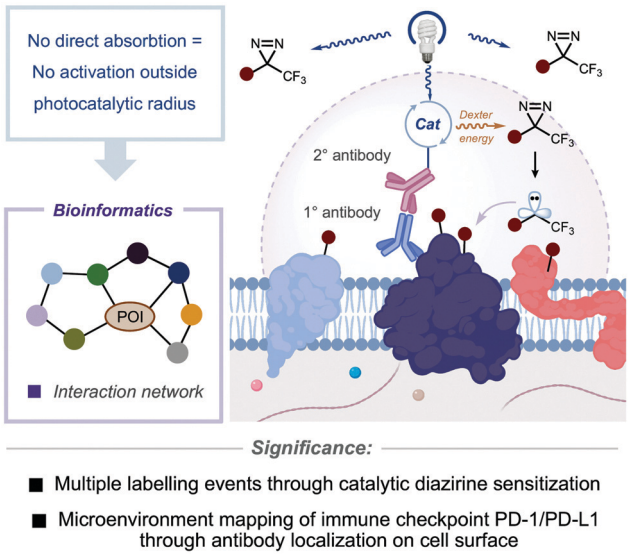


Fig. 21 MacMillan and co-workers generated interaction networks of cell surface proteins, including PD-L1, using their catalytic labeling methodology. Multiple tagging events enables the capture of low abundance membrane interactors. The data presented was higher resolution than analogous experiments using peroxidase-based methods, which can be attributed to the higher reactivity of the carbene intermediate.

for potential applications in immunotherapy where therapeutic efficacy is directly impacted by the contents of the immune synapse.

Generally, diazirines have seen a resurgence in the last few years as reagents for spatially restricted microenvironment elucidation. This increase in popularity may have arisen from their short solution half-life (in comparison to phenoxyl radicals and activated esters), which can increase the precision of the proteomic data, reducing the number of hits and therefore the time required for hit validation. The low cross-linking yields that diazirines suffer from may be offset by catalytic activation methodologies. This approach would present an empowering tool if new methodologies for its deployment are developed. Broadly, both stoichiometric and catalytic methods afford exquisite temporal control and are minimally disruptive to the cellular environment, presenting significant opportunities for those wishing to investigate dynamic signalling pathways or the effect of cellular stressors.

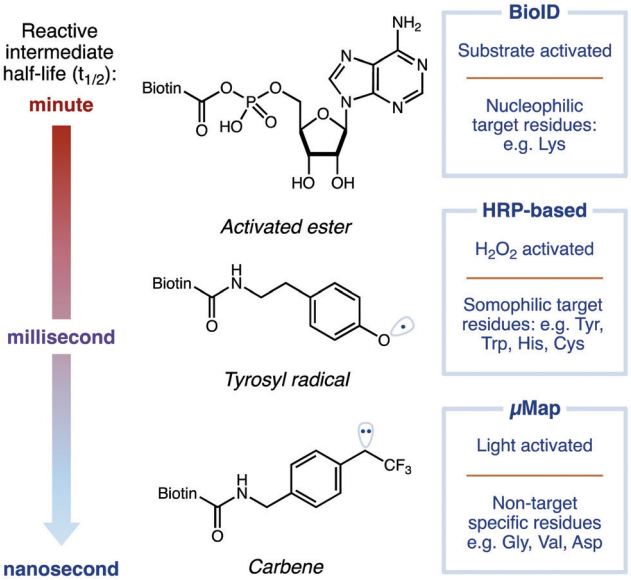


Fig. 22 Reactive intermediates commonly used for proximity labelling. The solution state half-life of the reactive intermediate determines the labelling radius. Different reactive intermediates target different protein residues.

6. Comparison of methods

In this review we have discussed a selection of methods designed for the elucidation of PPIs using high energy intermediates. Each class of intermediate possesses different properties that affect both the utility and scope of the labelling method. The greatest impact comes from the general reactivity of the intermediate, which can be expressed by the solution half-life ($t_{1/2}$). Broadly, the greater the solution half-life the larger the radius of the labelling method. Carbenes possess very short solution half-lives, so exhibit extremely short labelling radii, whereas activated esters have relatively long half-lives so can diffuse widely before cross-linking with proteins, leading to a large labelling radius (Fig. 22). Based upon this, users should take half-life into account when deciding which labelling technology to use to answer their desired biological question. It is also important to note that these proximity labelling approaches report on the distance from the "antenna" or bait, rather than providing conclusive PPIs. It is therefore crucial, especially with longer range technologies, to validate hits independently before assigning a new PPI.

7. Conclusions

The development of proximity labelling methodologies over the past 10 years has broadly impacted how the biomedical community studies cellular interactions. These methods have seen broad uptake, being used to examine protein interactions across many different biological systems both in a cellular context and *in vivo*. The transient and weak interactions that can be captured using these methods have the potential to unlock new therapeutic targets and help understand phenotype based clinical outcomes for existing therapeutics. Further development of new reactive intermediates has the potential to expand this field by allowing practitioners to vary the radius of labelling at will, thus capturing only the most relevant protein interactors. We hope that this review has highlighted that the union of chemistry and biology can present powerful tools that can impact human health and how we understand disease.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

A. D. T. would like to thank the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie Grant Agreement No. 891458.

Notes and references

1 R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. Maccoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young and B. Zhang, *Nat. Chem. Biol.*, 2018, **14**, 206–214.

2 H. Ruffner, A. Bauer and T. Bouwmeester, *Drug Discovery Today*, 2007, **12**, 709–716.

3 D. E. Scott, A. R. Bayly, C. Abell and J. Skidmore, *Nat. Rev. Drug Discovery*, 2016, **15**, 533–550.

4 G. Butland, J. M. Peregrín-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, M. Davey, J. Parkinson, J. Greenblatt and A. Emili, *Nature*, 2005, **433**, 531–537.

5 K. Kwon and D. Beckett, *Protein Sci.*, 2000, **9**, 1530–1539.

6 K. J. Roux, D. I. Kim, M. Raida and B. Burke, *J. Cell Biol.*, 2012, **196**, 801–810.

7 D. I. Kim, S. C. Jensen, K. A. Noble, B. Kc, K. H. Roux, K. Motamedchaboki and K. J. Roux, *Mol. Biol. Cell*, 2016, **27**, 1188–1196.

8 M. Ramanathan, K. Majzoub, D. S. Rao, P. H. Neela, B. J. Zarnegar, S. Mondal, J. G. Roth, H. Gai, J. R. Kovalski, Z. Siprashvili, T. D. Palmer, J. E. Carette and P. A. Khavari, *Nat. Methods*, 2018, **15**, 207–212.

9 T. C. Branon, J. A. Bosch, A. D. Sanchez, N. D. Udeshi, T. Svinkina, S. A. Carr, J. L. Feldman, N. Perrimon and A. Y. Ting, *Nat. Biotechnol.*, 2018, **36**, 880–898.

10 M. Kalkat, D. Resetca, C. Lourenco, P. K. Chan, Y. Wei, Y. J. Shiah, N. Vitkin, Y. Tong, M. Sunnerhagen, S. J. Done, P. C. Boutros, B. Raught and L. Z. Penn, *Mol. Cell*, 2018, **72**, 836–848.

11 C. V. Dang, *Cell*, 2012, **149**, 22–35.

12 K. F. Cho, T. C. Branon, S. Rajeev, T. Svinkina, N. D. Udeshi, T. Thoudam, C. Kwak, H. W. Rhee, I. K. Lee, S. A. Carr and A. Y. Ting, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 12143–12154.

13 R. Varnaitė and S. A. MacNeill, *Proteomics*, 2016, **16**, 2503–2518.

14 L. Trinkle-Mulcahy, *F1000Research*, 2019, **8**, 135.

15 J. N. Rodríguez-López, D. J. Lowe, J. Hernández-Ruiz, A. N. P. Hiner, F. García-Cánovas and R. N. F. Thorneley, *J. Am. Chem. Soc.*, 2001, **123**, 11838–11847.

16 S. Sato and H. Nakamura, *Molecules*, 2019, **24**, 3890.

17 N. Kotani, J. Gu, T. Isaji, K. Udaka, N. Taniguchi and K. Honke, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7405–7409.

18 B. A. DeGraff, D. W. Gillespie and R. J. Sundberg, *J. Am. Chem. Soc.*, 1974, **96**, 7491–7496.

19 X. W. Li, J. S. Rees, P. Xue, H. Zhang, S. W. Hamaia, B. Sanderson, P. E. Funk, R. W. Farndale, K. S. Lilley, S. Perrett and A. P. Jackson, *J. Biol. Chem.*, 2014, **289**, 14434–14447.

20 H.-W. Rhee, P. Zou, N. D. Udeshi, J. D. Martell, V. K. Mootha, S. A. Carr and A. Y. Ting, *Science*, 2013, **339**, 1328–1331.

21 J. D. Martell, T. J. Deerinck, Y. Sancak, T. L. Poulos, V. K. Mootha, G. E. Sosinsky, M. H. Ellisman and A. Y. Ting, *Nat. Biotechnol.*, 2012, **30**, 1143–1148.

22 S. S. Lam, J. D. Martell, K. J. Kamer, T. J. Deerinck, M. H. Ellisman, V. K. Mootha and A. Y. Ting, *Nat. Methods*, 2014, **12**, 51–54.

23 J. Paek, M. Kalocsay, D. P. Staus, L. Wingler, R. Pascolutti, J. A. Paulo, S. P. Gygi and A. C. Kruse, *Cell*, 2017, **169**, 338–349.

24 B. T. Lobingier, R. Hüttenhain, K. Eichel, K. B. Miller, A. Y. Ting, M. von Zastrow and N. J. Krogan, *Cell*, 2017, **169**, 350–360.

25 S. A. Myers, J. Wright, R. Peckner, B. T. Kalish, F. Zhang and S. A. Carr, *Nat. Methods*, 2018, **15**, 437–439.

26 X. D. Gao, L.-C. Tu, A. Mir, T. Rodriguez, Y. Ding, J. Leszyk, J. Dekker, S. A. Shaffer, L. J. Zhu, S. A. Wolfe and E. J. Sontheimer, *Nat. Methods*, 2018, **15**, 433–436.

27 F. M. Fazal, S. Han, K. R. Parker, P. Kaewsapsak, J. Xu, A. N. Boettiger, H. Y. Chang and A. Y. Ting, *Cell*, 2019, **178**, 473–490.

28 Y. Zhou, G. Wang, P. Wang, Z. Li, T. Yue, J. Wang and P. Zou, *Angew. Chem., Int. Ed.*, 2019, **131**, 11889–11893.

29 D. A. Fancy and T. Kodadek, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6020–6024.

30 S. Sato, K. Morita and H. Nakamura, *Bioconjugate Chem.*, 2015, **26**, 250–256.

31 K. Shiraiwa, R. Cheng, H. Nonaka, T. Tamura and I. Hamachi, *Cell Chem. Biol.*, 2020, **27**, 970–985.

32 R. A. G. Smith and J. R. Knowles, *J. Am. Chem. Soc.*, 1973, **95**, 5072–5073.

33 J. Wang, G. Burdzinski, J. Kubicki, M. S. Platz, R. A. Moss, X. Fu, P. Piotrowiak and M. Myahkostupov, *J. Am. Chem. Soc.*, 2006, **128**, 16446–16447.

34 K. A. Jacobson, *Anal. Biochem.*, 1983, **140**, 8–24.

35 M. Vila-Perellö, M. R. Pratt, F. Tulin and T. W. Muir, *J. Am. Chem. Soc.*, 2007, **129**, 8068–8069.

36 M. Suchanek, A. Radzikowska and C. Thiele, *Nat. Methods*, 2005, **2**, 261–267.

37 H. W. Ai, W. Shen, A. Sagi, P. R. Chen and P. G. Schultz, *ChemBioChem*, 2011, **12**, 1854–1857.

38 R. E. Kleiner, L. E. Hang, K. R. Molloy, B. T. Chait and T. M. Kapoor, *Cell Chem. Biol.*, 2018, **25**, 110–120.
39 Y. Yang, H. Song, D. He, S. Zhang, S. Dai, X. Xie, S. Lin, Z. Hao, H. Zheng and P. R. Chen, *Nat. Protoc.*, 2017, **12**, 2147–2168.
40 A. Sinz, *Angew. Chem., Int. Ed.*, 2018, **57**, 6390–6396.
41 F. Liu, D. T. S. Rijkers, H. Post and A. J. R. Heck, *Nat. Methods*, 2015, **12**, 1179–1184.
42 A. J. Burton, M. Haugbro, E. Parisi and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 12041–12049.
43 N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446–461.
44 A. J. Burton, M. Haugbro, L. A. Gates, J. D. Bagert, C. D. Allis and T. W. Muir, *Nat. Chem.*, 2020, **12**, 520–527.
45 E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007–2010.
46 Y. Tanaka and J. J. Kohler, *J. Am. Chem. Soc.*, 2008, **130**, 3278–3279.
47 J. E. McCombs, C. Zou, R. B. Parker, C. W. Cairo and J. J. Kohler, *ACS Chem. Biol.*, 2016, **11**, 185–192.
48 D. C. McCutcheon, G. Lee, A. Carlos, J. E. Montgomery and R. E. Moellering, *J. Am. Chem. Soc.*, 2020, **142**, 146–153.
49 J. B. Geri, J. V. Oakley, T. Reyes-Robles, T. Wang, S. J. McCarver, C. H. White, F. P. Rodriguez-Rivera, D. L. Parker, E. C. Hett, O. O. Fadeyi, R. C. Oslund and D. W. C. MacMillan, *Science*, 2020, **367**, 1091–1097.
50 N. J. Turro, Y. Cha, I. R. Gould and R. A. Moss, *J. Photochem.*, 1987, **37**, 81–86.

