

This is a repository copy of *Tag thy neighbour: nanometre-scale insights into kinetoplastid parasites with proximity dependent biotinylation*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/185631/>

Version: Accepted Version

Article:

Geoghegan, Vincent, Mottram, Jeremy Charles orcid.org/0000-0001-5574-3766 and Jones, Nathaniel orcid.org/0000-0001-7328-4487 (2022) Tag thy neighbour: nanometre-scale insights into kinetoplastid parasites with proximity dependent biotinylation. *Frontiers in cellular and infection microbiology*. ISSN: 2235-2988

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Tag thy neighbour: nanometre-scale insights into kinetoplastid parasites with proximity dependent biotinylation

Vincent Geoghegan, Jeremy C. Mottram, Nathaniel G. Jones

York Biomedical Research Institute and Department of Biology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK

Abstract

Proximity labelling is a powerful and rapidly developing technology for exploring the interaction space and molecular environment of a protein of interest at the nanometre scale. In proximity labelling, a promiscuous biotinylating enzyme is genetically fused to the protein of interest, initiation of labelling then results in the biotinylating enzyme generating reactive biotin which covalently 'tags' nearby molecules. Importantly, this labelling takes place *in vivo* whilst the protein of interest continues to perform its normal functions in the cell. Due to its unique advantageous characteristics, proximity labelling is driving discoveries in an ever increasing range of organisms. Here, we highlight the applications of proximity labelling to the study of kinetoplastids, a group of eukaryotic protozoa that includes trypanosomes and *Leishmania* which can cause serious disease in humans and livestock. We first provide a general overview of the proximity labelling experimental workflow including key labelling enzymes used, proper experimental design with appropriate controls and robust statistical analysis to maximise the amount of reliable spatial information that is generated. We discuss studies employing proximity labelling in kinetoplastid parasites to illustrate how these key principles of experimental design are applied. Finally, we highlight emerging trends in the development of proximity labelling methodology.

Introduction

Discovery of interaction partners is often a key step towards understanding the biological role of a protein of interest. A common approach is affinity purification of the protein of interest followed by identification of co-purifying proteins by mass spectrometry (AP-MS). AP-MS is a well established technique that can yield valuable insights (Gingras et al., 2007). However, the success of an AP-MS experiment depends on finding lysis conditions that are sufficient to solubilise the protein of interest (bait), but not so stringent that associations with interactors (prey) are disrupted. Preserving interactions therefore necessitates relatively gentle lysis and washing conditions, unfortunately this can be permissive to formation of post-lysis interactions which do not occur *in vivo* (Branon et al., 2017). Proximity labelling is a new paradigm for exploring interactions involving a protein of interest which bypasses the requirement to preserve interactions during lysis and enrichment. This key advantage of proximity labelling lies in the fact that the protein of interest is expressed as a fusion to an enzyme that generates reactive biotin molecules which then covalently tag nearby proteins *in vivo* (Samavarchi-Tehrani et al., 2020). Interacting proteins will continue to associate and dissociate during the labelling period, potentially leading to an accumulation of biotinylation signal over time, in contrast to AP-MS which can only capture interactors at the point of lysis. Two hybrid interaction assays such as yeast two hybrid, like proximity labelling, can detect *in vivo* interactions and potentially offer a very detailed view of interaction networks by virtue of the pairwise screening for bait-prey interactions. In two hybrid assays, direct interaction partners can be discovered in a high-throughput manner, although the set up costs can be quite high and proteins are often expressed outside of their normal context (Mehla et al., 2017).

Proximity labelling is a broad term collecting together a range of approaches which can be used to characterise proximal molecular space not just for proteins, but other biomolecules such as DNA and RNA. Since the majority of proteomic studies use a bait fused to an enzyme which generates reactive

biotin to carry out proximity dependent biotinylation, we focus on this subset of proximity labelling here. An overview of the complete proximity biotinylation workflow is shown in **Fig. 1**.

The concept of proximity dependent biotinylation was proposed by Choi-Ree et al. based on information gained from prior fundamental studies on the structure and function of *E. coli* biotin ligase (Choi-Rhee et al., 2004). *E. coli* BirA biotin ligase specifically biotinylates a single lysine in biotin carboxyl carrier protein in a two step reaction involving a reactive intermediate that is normally held in the active site. However, previous work identified two mutant forms of BirA, G115S and R118G which have reduced affinity for the biotinylation reaction intermediate biotin adenylate (biotinyl-5'-AMP) compared to the wild-type enzyme (Kwon and Beckett, 2000). Since biotinyl-5'-AMP is a reactive and short lived biotin adenylate, this led Choi-Ree et al. to propose that "*these mutant proteins might leak bio-5'-AMP into the solvent where it would act as a promiscuous chemical biotinylation reagent in a proximity-dependent manner*" (Choi-Rhee et al., 2004) (**Fig. 2A**). Indeed, *in vitro* experiments were performed demonstrating proximity dependent biotinylation using the R118G promiscuous biotin ligase variant, BirA*. *In vivo* experiments were also performed showing the promiscuous nature of biotinylation, but targets were not identified. The first demonstration of proximity dependent biotinylation *in vivo* arrived much later in a study by Roux et al. who fused BirA* (R118G) to lamin-A and demonstrated enrichment of proximal and interacting proteins, naming the technique BioID (Roux et al., 2012).

Several excellent general overviews of proximity biotinylation have been written which should be consulted when planning experiments (Samavarchi-Tehrani et al., 2020; Qin et al., 2021a). Here, we provide an overview of the technique and focus on its practical application to research on kinetoplastids.

Biotin ligases and peroxidases

There are two main classes of enzyme used for proximity dependent biotinylation, promiscuous biotin ligases and peroxidases (Samavarchi-Tehrani et al., 2020). The original *E. coli* R118G biotin ligase mutant has been widely used in over 300 studies and remains an important tool. Labelling kinetics of BirA* are relatively slow, requiring 18-24 hr incubations with exogenous biotin to achieve optimal labelling, which does not allow studies of more dynamic processes that occur in seconds to minutes. However, slow kinetics may help to reduce background as only true interactors will dwell long enough to become biotinylated, which may explain the success of BirA* for characterising stable protein complexes (Gingras et al., 2019). Although *in vivo* labelling by BirA* is relatively slow, it is worth noting that first permeabilising cells and performing '*in vitro* BioID' enables labelling in 10-15mins (Remnant et al., 2019). This suggests that BirA* can perform rapid labelling under certain conditions, in this case possibly due to effects of detergents or the removal of quenchers. Further investigation of factors influencing promiscuous biotinylation is therefore required, as our understanding of the mechanism is incomplete.

More recent engineered biotin ligases have been developed such as BioID2, BASU, AirlID, ultraID, miniTurboID and TurboID (Kim et al., 2016; Branon et al., 2018; Ramanathan et al., 2018; Kido et al., 2020; Zhao et al., 2021). These enzymes show a spectrum of labelling kinetics with BioID2 having modest improvements in labelling efficiency over BirA*, whilst AirlID enables faster labelling in ~3hrs. TurboID and miniTurboID were created by directed evolution of BirA* and rapidly biotinylate proximal proteins in 10-15mins (Branon et al., 2018). Because activity screening for these variants was carried out at the growth temperature of yeast, 26°C, they may be well suited for use in organisms that are grown at lower temperatures such as *Leishmania* promastigotes. They have already been used in a number of studies, however they each have limitations. TurboID, which is the more efficient enzyme of the two, also has a higher affinity for biotin and therefore carries out extensive biotinylation using the low levels of biotin already present in culture media. This prevents temporal control of biotinylation, thus decreasing spatial resolution and increasing background, because the enzyme is constitutively

biotinylating proximal proteins as soon as the transgene is expressed. MiniTurboID labels more slowly but has less background activity in the absence of exogenous biotin, enabling better temporal control of labelling (Branon et al., 2018). The DNA binding domain present in BirA was deleted to create miniTurbo which may help to reduce background labelling of chromatin associated proteins, however this appears to have destabilised the protein in some contexts (May et al., 2020). BASU, an engineered biotin ligase from *Bacillus subtilis* also lacks a DNA binding domain and was presented as enabling proximity labelling in a timescale of 1min, however a later study did not find it was significantly faster than BirA* (Branon et al., 2018). Most recently, ultraID has been developed by directed evolution of biotin ligase from the thermophilic bacterium *Aquifex aeolicus* and appears to be a very promising enzyme, combining fast labelling kinetics (10mins), small size (20kDa) and low background activity in the absence of exogenous biotin (Zhao et al., 2021). The directed evolution study leading to ultraID showed that just two mutations in the biotin binding site generated a highly active promiscuous biotin ligase, contrasting with the directed evolution of TurboID which has a total of 15 mutations compared to the wild-type enzyme. Aside from the R118S mutation, the mutations in TurboID appear to be spread throughout the structure of the biotin ligase, exactly how they contribute to augmenting promiscuous biotinylation activity is unclear, further highlighting the need for fundamental studies on the mechanism of proximity dependent biotinylating enzymes.

The focus of biotin ligase engineering efforts has been to produce enzymes with faster labelling kinetics. This is evidently required for studies of dynamic cellular processes, but faster is not necessarily better for every application. A side by side comparison of nuclear pore proteins Nup43 and Nup53 fused to BirA* or TurboID in mammalian cells, showed that coverage of the nuclear pore complex was similar for both biotinylating enzymes but that the total number of proteins identified in the TurboID samples was ~2.7 fold higher, the additional hits comprising proteins present in other subcellular compartments such as the ER and mitochondria (May et al., 2020). This indicates the higher activity of TurboID may lead to a higher background due to off target labelling of bystander proteins, whereas the slower labelling of BirA* may actually lead to a cleaner background because only true interactors reside long enough within the labelling radius to become biotinylated.

Peroxidase based proximity biotinylation proceeds via a different mechanism compared to biotin ligase based approaches. Cells expressing the peroxidase fused to a protein of interest are first incubated with a phenol derivative of biotin such as biotin-phenol. Proximity biotinylation is then initiated with the addition of H₂O₂. Peroxidases such as APEX2 catalyse H₂O₂ mediated oxidation of biotin-phenol, generating highly reactive phenoxyl-biotin radicals which then covalently tag electron rich amino acid side chains, such as tyrosine, in proximal proteins (Lam et al., 2015). A key advantage of APEX2 catalysed proximity biotinylation is that it is very rapid, permitting labelling times as short as 1 min to be used which has allowed dynamic processes such as G-protein coupled receptor signalling to be followed (Lobingier et al., 2017; Paek et al., 2017). If the protein of interest is a cell surface protein, horseradish peroxidase (HRP) can be used to biotinylate proximal proteins either by expressing the protein fused to HRP (Miyagawa-Yamaguchi et al., 2014), or targeting the protein with a specific HRP-conjugated antibody (Bar et al., 2018). HRP has a higher activity than APEX2, unfortunately it is not active in the cytosol which limits its application to extracellular facing proteins (Hopkins et al., 2000). While being well suited to taking rapid snapshots of proximal molecular environments, peroxidase based labelling has its own limitations. Labelling is initiated by H₂O₂ which is toxic and can lead to membrane disruption, thus it may be less suitable for whole organism proximity biotinylation studies. Finally, as well as the expressed peroxidase fusion protein, endogenous peroxidases may also catalyse biotinylation and contribute to the background signal.

Cellular compartments such as the nucleus or endoplasmic reticulum (ER) each have their own distinct chemical micro-environments that can strongly influence the activity of biotinylating enzymes. As part of their thorough characterization of evolved BirA variants, Branon *et al.* compared the activity of BirA*, TurboID and miniTurbo in the cytosol, nucleus, ER lumen, ER membrane and mitochondrial matrix of HEK293T cells (Branon et al., 2018). They observed that all 3 ligases had lower activity in

the ER lumen and mitochondrial matrix compared to the other compartments. Comparing between the three ligase variants, BirA* generated a relatively consistent biotinylation signal across all compartments, whilst the activity of TurboID was quite variable, being lowest in the nucleus, but higher in the ER lumen and mitochondrial matrix, relative to BirA*. In the cytosol, TurboID and miniTurbo had highest activity relative to BirA*, compared to the other organelles.

The labelling radius

A key question often raised when discussing proximity biotinylation techniques is the labelling radius. The first estimation of the *in vivo* labelling radius used the Y-complex, a sub-complex of the nuclear pore complex, as a molecular ruler to measure the labelling distance (Kim et al., 2014). BirA* fusions were created for 5 members of the Y-complex which has known architecture and dimensions. Analysis of the detected abundance of co-enriched subunits indicated the labelling radius of BirA* was ~10nm. In a separate study, BioID was performed in cells expressing BirA* fused to a centromeric histone variant (Remnant et al., 2019). Deposition of biotin along chromatin was directly visualised by electron microscopy which revealed a narrow distribution either side of CENP-A BirA* along the linear chromatin, with a radius of ~10nm, in agreement with the study on the Y-complex. For context, an averaged sized protein of 50kDa has a diameter of ~5nm assuming a spherical shape (Erickson, 2009), therefore proximity biotinylation has a high spatial resolution; for investigations of large macromolecular complexes multiple subunits may need to be fused to a biotinylator or a cross-linking step included to improve coverage (Geoghegan et al., 2021). However, as has been previously highlighted, it is more accurate to view the labelling radius as a contour map, not having a hard proximity cutoff, but rather a decreasing probability of labelling with distance (Qin et al., 2021a). The exception to this are cases when the biotinylator is close to a membrane which does act as a hard cutoff, being impermeable to reactive biotin species.

It is also important to consider the nature of the concentration gradient of reactive biotin molecules formed from the source. A gradient will form as a result of diffusion away from the source but also due to quenching/breakdown of the reactive biotin molecules over time. Both these processes are expected to combine to create an *exponentially* decreasing concentration of reactive biotin (**Fig. 2B**) as a function of distance from the biotinylator which would contribute to the high spatial resolution (Wartlick et al., 2009). Thus biotinylation of a complex will drop off sharply with distance from the bait, direct interactors will be strongly labelled, but indirect interactors much less so, this is indeed what is observed from proximity biotinylation data. On this basis, it is tempting to infer distance of a proximal protein from the bait by analysing its relative amounts as quantified by mass spectrometry in the enriched biotinylated material. However, there are too many factors other than distance which determine the extent of proximity biotinylation, such as availability of lysine residues, residence time in the complex, orientation of the protein and copy number of the protein.

Measurements of the labelling radius of proximity biotinylation using BirA* indicate a high spatial resolution of ~10nm, however the labelling radius will be determined in each case by the half life of the reactive intermediate involved and the subcellular environment. Also, measurements of labelling radius are based on stable, long lived structures. In reality, there will be many dynamic processes taking place in the vicinity, proteins may diffuse into the labelling radius and become biotinylated without being part of the structure. Additionally, for conventionally designed studies, the labelling radius is active throughout the entire lifetime of the protein, from its point of synthesis to the point of degradation. This can make the resulting datasets quite complex and could be a particular issue for the newer generation of rapid biotin ligases which may be more prone to tagging 'bystander' proteins. An excellent control for off-target biotinylation is a spatial reference, where a cell line is generated expressing the biotinylator localised to the same sub-cellular region as the protein of interest. The control cell line could even express a protein of potential interest tagged with a biotinylator. In this scenario, the two tagged baits act as reciprocal controls (**Fig. 2C** and see experimental design section).

Proximity biotinylation in kinetoplastids

Kinetoplastids are a large group of flagellated protozoa named after the kinetoplast, a characteristic mass of DNA contained in the single mitochondrion (Povelones, 2014). Some kinetoplastids exist as free-living organisms, whilst others require a host to survive and can infect a range of organisms such as plants, insects or animals (Kostygov et al., 2021). Two genera of kinetoplastids cause serious diseases in humans and livestock: *Leishmania* causing the leishmaniases and *Trypanosoma* causing sleeping sickness and Chagas disease (Burza et al., 2018; Libisch et al., 2021; Pays and Nolan, 2021). As well as for their impact on human health, kinetoplastids are also studied for their rather unique way among eukaryotes of carrying out molecular processes, such as absence of introns, constitutive transcription of genes, widespread trans-splicing and highly divergent or even unique protein complexes of fundamental importance to cell biology, such as the kinetochore or pre-replication complex (Tiengwe et al., 2012; Akiyoshi and Gull, 2014; Clayton, 2016; De Pablos et al., 2016). Comprehensive molecular toolkits allowing functional genetics make these organisms powerful non-opisthokont model organisms.

A number of studies (**Table 1**) have made use of proximity biotinylation to investigate protein complexes in different sub-cellular compartments of kinetoplastids, with a notable focus on characterising cytoskeletal associated complexes, for which proximity biotinylation is well suited. We highlight selected studies that illustrate the range of technical approaches used in kinetoplastids.

Trypanosoma brucei

The majority of proximity biotinylation studies in kinetoplastids to date have investigated protein complexes in procyclic form *Trypanosoma brucei* (**Fig. 3A**). This is likely because relative to other kinetoplastids, a well developed genetic toolbox in *T. brucei* has been available for longer, making it straightforward to endogenously tag a protein of interest with a promiscuous biotin ligase or overexpress a bait protein fused to a biotin ligase. Morriswood *et al.* reported the first successful use of proximity biotinylation in kinetoplastids in 2013, in fact this was only the second reported use of proximity biotinylation to probe *in vivo* complexes after the original Roux *et al.* study (Morriswood et al., 2013). The authors investigated a discrete cytoskeletal structure known as the bilobe, which localises to just above the flagellar pocket collar. To interrogate the composition of the bilobe, a known marker, MORN1, was N-terminally tagged with BirA* and this construct was expressed from the ribosomal locus under the control of the tetracycline inducible promoter, giving modest overexpression compared to endogenous levels. Correct expression and localization of the BirA*-MORN1 was first confirmed by western blotting and fluorescence microscopy. Biotinylation of proximal proteins was carried out for 24hrs, before insoluble cytoskeletal proteins were extracted to provide some enrichment for the complex of interest. Biotinylated material was then purified from SDS solubilised cytoskeletal extract. Three biological replicates were performed, no tetracycline induction of BirA*-MORN1 expression served as the control group. Protein hits were inspected manually to determine candidate proximals, those that had a high number of identified peptides in the induced vs. non-induced samples and with an uncharacterised localisation were selected for screening by fluorescence microscopy. Of the 10 screened proteins, 7 were found to be novel components of the bilobe structure, demonstrating the power of proximity biotinylation for discovering cytoskeletal protein complexes that are otherwise biochemically challenging to profile due to their low solubility.

Subsequent applications of proximity biotinylation in *T. brucei* have mostly been very similar in experimental design. The majority of studies relied on wild-type parasites or uninduced parasites as negative controls, rather than using spatial references to estimate off target biotinylation. The exception to this is work by De Pablos *et al.* and Pyrih *et al.*, where compartment matched spatial references were used to more accurately estimate off target biotinylation (De Pablos et al., 2017; Pyrih et al., 2020). For example, Pyrih *et al.* investigated 4 mitochondrial bait proteins by proximity biotinylation, using the baits as reciprocal controls.

Overall, 13 of the 16 studies used BirA* tagged proteins and these were mainly overexpressed under the control of a tetracycline inducible promoter. Overexpression has the potential for increasing off-target biotinylation, but this is mitigated by adjusting expression levels of the fusion protein with tetracycline to approximately match levels of the endogenous protein. Dong *et al.* endogenously tagged Spef1 with BirA* to characterise microtubule quartet associated proteins residing in the region of the flagellum between the bilobe and basal body (Dong et al., 2020). Correct localization of the fusion protein was verified by microscopy and spatially focused biotinylation was demonstrated by incubating parasites with biotin, followed by visualization of proximity biotinylation with fluorescent streptavidin. To maximise capture of proximal proteins, separate experiments were performed with both N and C-terminally BirA* tagged Spef1. As with previous studies examining cytoskeletal complexes, a pre-enrichment step was carried out by isolating the cytoskeleton which remains in the insoluble fraction following lysis with a gentle detergent such as NP-40. Unusually for this type of study, an estimated absolute amount was calculated for each identified protein based on the exponentially modified protein abundance index (emPAI) (Ishihama et al., 2005). Putative proximals present in high absolute amounts and absent in the control samples were selected for follow up validation and functional studies. Here control samples were wild-type parasites. All 8 of the uncharacterised proximal proteins followed up were found to either co-localise with Spef1 or to nearby flagellar associated structures. One of these, SAF1, was shown to be a novel component of the basal body and required for correct spatial organization of structures at the base of the flagellum.

Another example of exploring proximal proteins by using both N and C terminal BirA*-bait fusions are two studies by Zhou *et al.*, in which CIF1 proximal proteins were identified (Zhou et al., 2016, 2018a). CIF1 localises to the tip of the flagellum attachment zone and has an essential role in cytokinesis initiation. CIF1-BirA* was overexpressed in the first study and proximity biotinylation identified 160 potential proximal proteins, although this number appears to be based on simple manual inspection of identified proteins. Twelve of these proteins were imaged by fluorescence microscopy, 4 of which were found to be close to CIF1, one of these was named CIF2 and association with CIF1 was confirmed by co-immunoprecipitation. A third member of the putative complex, CIF3 was discovered in separate work through a tagging and fluorescence microscopy screen. Because CIF3 was not in the initial proximity biotinylation screen using a C-terminally tagged bait, further work was carried out using N-terminally BirA* tagged CIF1 as a bait. This second proximity biotinylation analysis of CIF1 identified more than 500 proteins in total, proteins deemed to be common contaminants were manually removed from the dataset leaving 305 putative proximals. Fluorescence microscopy localisation screening reduced this down to 52 proteins confirmed as proximal to CIF1, 33 of which were detected in both the BirA*-CIF1 and CIF1-BirA* datasets, showing good, but not complete overlap between datasets generated from the different tagged termini.

There are few reports comparing proximity biotinylation with other approaches commonly used for identifying interaction partners. Baudouin *et al.* compared the performance of proximity biotinylation, immunoprecipitation and yeast two hybrid assay in *T. brucei* for detecting proteins interacting with TAC102, a mitochondrial component of the tripartite attachment complex (TAC) which links the basal body of the flagellum to mitochondrial DNA (kDNA) (Baudouin et al., 2020). Proteins are usually targeted to the mitochondrion by an N-terminal targeting signal, which would normally preclude fusing the BirA* ligase here, however in the case of TAC102 the mitochondrial targeting signal is actually located at the C-terminus, therefore BirA* was fused to the N-terminus and correct mitochondrial import of the fusion protein was observed, an excellent example of where correct placement of the BirA* ligase on the bait was critical to the success of the project. In this work, 4 biological replicates were used and enrichment was calculated relative to uninduced control samples. To define proximal proteins, a t-test was performed, proteins enriched at least 3-fold and with significance $p \leq 0.01$ were considered proximal, giving a final list of 77 proteins, 47 of these were predicted to be mitochondrial. From the immunoprecipitation experiment of native untagged TAC102, 100 proteins were found significantly enriched, 49 of which were predicted to be mitochondrial. Only 9 proteins were detected as significantly enriched in both proximity biotinylation and immunoprecipitation approaches, of which

3 localised to the kDNA/TAC region. The yeast two-hybrid screen identified a known TAC component, p166, but otherwise generated few confident hits.

In this comparison, results from proximity biotinylation were somewhat disappointing because using BirA*-TAC102 as bait did not lead to detection of any other known TAC proteins. Immunoprecipitation results were slightly better, identifying one known TAC component, p166. The two approaches did identify 3 potential novel TAC components, but the low recall of known complex members suggests that TAC presents distinct challenges and further optimisation is required. Biotinylation was carried out over 6hrs, so one possible improvement could be extending the duration of biotinylation.

In order to capture snapshots of the local protein environment with higher temporal resolution, proximity biotinylating methods using enzymes with faster kinetics need to be established in kinetoplastids. To this end, Velez-Ramirez *et al.* reported the first use of APEX2 in kinetoplastids, using APEX2 fused to flagellar proteins to explore flagellum subdomains in *T. brucei* (Vélez-Ramírez *et al.*, 2021). Three flagellar localised baits were used as spatial probes, DRC1, AC1 and FS179, all fused with APEX2 at their C-terminus by endogenous tagging. DRC1 is present at the axoneme, an arrangement of microtubules forming the core structure of the flagellum, AC1 localises to the tip of the flagellum and FS179 localises to the flagellum attachment zone and is excluded from the tip. Procyclic parasites were pre-incubated with biotin-phenol for 1hr, then labelling was carried out for 1min by treatment with H₂O₂, followed by a quenching step. Parasites were lysed under relatively gentle conditions so that the lysate could be fractionated into soluble and non-soluble parts. The non-soluble fraction containing the majority of the cytoskeletal proteins was later solubilised in SDS for capture of biotinylated material. Very stringent and extensive streptavidin bead washing conditions were used in this workflow, for example buffers containing 8M urea, taking full advantage of the strong biotin-streptavidin interaction to strip the beads of non-biotinylated proteins. To define a proximal proteome for the baits, a semi-quantitative strategy was used based on spectral counting. For the bait AC1, known flagellum tip proteins in the dataset were used to set enrichment thresholds, proteins enriched above this threshold relative to samples from wild-type parasites were considered proximal. The AC1 proximity proteome, at 48 proteins, was more focussed than the 697 proteins in the DRC1 proximity proteome and 400 in the FS179 proximity proteome, consistent with more spatially restricted labelling. Interestingly, proteins with possible roles in signal transduction at the flagellum tip, such as adenylate cyclases, were detected, pointing towards future applications of the methodology in kinetoplastids to examine much more dynamic processes than flagellum structure.

Leishmania mexicana

Proximity biotinylation has only relatively recently been applied to study *Leishmania* cell biology with a few examples reported to date (**Fig. 3B**). Kelly *et al.* were the first to report the use of proximity biotinylation in *Leishmania*, using the approach to explore proteins proximal to KHARON (KH) (Kelly *et al.*, 2020). KH is a cytoskeletal protein important for replication of the intracellular amastigote stage with multiple localisations: at the base of the flagellum, the subpellicular microtubules around the cytosol and mitotic spindles. To discover KH proximal proteins that may be important for function at the subcellular regions, BirA*-KH was episomally overexpressed and biotinylation carried out overnight. The cytoskeletal fraction was purified for subsequent streptavidin enrichment of biotinylated material. The overall workflow was quite straightforward, with 2 biological replicates used and comparing to wild-type parasites to determine proximal proteins. Spectral counts for proteins in wild-type parasites were simply subtracted from the spectral counts recorded for BirA*-KH expressing parasites. Ribosomal proteins were manually removed as likely contaminants, hits with a difference in spectral counts >10 were then considered proximal, although a basal body protein of potential interest was included despite not meeting the criteria. No statistical testing was performed, but to provide additional evidence of interactors a tandem affinity purification (TAP) experiment on KH was carried out. KH is associated with insoluble cytoskeletal structures, not an issue for proximity biotinylation but a challenge for an immunoprecipitation type approach where interactions generally need to be

preserved during purification. To provide a route forward, the authors endogenously N-terminally tagged KH with the HBH tag. This tag contains a bacteria derived peptide that is biotinylated by endogenous biotin ligases in yeast and mammals, flanked by 6xHis, enabling TAP enrichment under denaturing conditions (Tagwerker et al., 2006). Parasites were cross-linked with formaldehyde to covalently link associating proteins, then lysed in 8M urea to solubilise complexes before denaturing TAP enrichment of KH associated complexes. Material was TMT labelled for a fully quantitative approach. Comparing the proximity biotinylation and TAP results revealed 3 proteins detected by both strategies, two of which were validated as novel KH associated proteins and demonstrated to be required for amastigote proliferation in macrophages.

Our group has recently completed two projects where we extended the use of proximity biotinylation to understanding fundamental aspects of chromatin biology in *Leishmania* (Geoghegan et al., 2021; Jones et al., 2021). We firstly discovered that we could increase the capture of proximal proteins by performing a limiting DSP cross-linking step after proximity biotinylation, an approach we termed XL-BioID (Geoghegan et al., 2021). This led to a ~10 fold increase in proximal protein signal intensity. We applied this technique to protein kinases belonging to the *Leishmania* kinetochore, a large complex that assembles on chromosomes, linking them to the mitotic spindle for segregation during mitosis. Kinetochore kinases KKT2, KKT3 and KKT19 were endogenously tagged with BirA* and used as baits in the XL-BioID experiment to explore the *Leishmania* kinetochore. We could detect 16 of the 25 predicted inner kinetochore proteins and discovered two novel kinetochore proteins, KKT24 and KKT26. Perhaps an underappreciated feature of proximity biotinylation is that the data it provides is not just a parts list for a complex of interest, rather it is also a snapshot of the dynamic cellular processes taking place in the vicinity of the bait protein. This aspect of proximity biotinylation data was apparent from our kinetochore study where we detected factors important for regulating chromatin structure, chromosome segregation factors and DNA replication factors proximal to kinetochore kinases. Interestingly, several nucleolar proteins were found in our proximity data, suggesting that the kinetochores associate with the nucleolus during the cell cycle which may be important for regulating their assembly.

To gain insight into assembly of the kinetochore during the cell cycle we used miniTurbo tagged KKT3, allowing us to take 30 minute biotinylation snapshots of the kinetochore in hydroxyurea synchronised parasites at G1/S, S and G2/M phases. Coverage of core kinetochore proteins was lower compared to using BirA*, for example at G1/S, 6 proximal KKTs were detected. Possible reasons for this could be due to different placement of the tag on the bait and faster labelling kinetics generating more noise. Nonetheless, we could follow changes in abundance amongst the detected proteins. For example, KKT4 was weakly proximal at G1/S, but became significantly more enriched during the cell cycle, reflecting its possible role in linking the kinetochore to the microtubule spindle, which takes place later in the cell cycle (Llauró et al., 2018).

Taking advantage of the increased capture of proximal proteins by XL-BioID in our investigation of the *leishmania* kinetochore (Geoghegan et al. 2021), we developed proximity phosphoproteomics; enriching phosphopeptides from proximal proteins to add another dimension to proximity biotinylation data, namely the ability to track individual phosphosites at a complex of interest (Geoghegan et al. 2021). We quantified proximal phosphosites at the kinetochore as parasites progressed through the cell cycle. Label free quantification was used for both proximal proteins and phosphosites. Since the proximal phosphosites were low in abundance and likely to have higher variability, we used 5 biological replicates to give us sufficient statistical power to determine proximal phosphosites with limma, comparing intensities with a spatial reference. Limma is an R package for statistical analysis of various omics data, its application to proximity biotinylation data is discussed in more detail below (Ritchie et al. 2015). Our analysis revealed groups of phosphorylation sites with distinct dynamics, some changed in abundance in the same direction as the parent protein, but others showed highly dynamic behaviour during the cell cycle which was not a straightforward reflection of abundance of the protein at the kinetochore. These latter proximal phosphosites, such as KKT2 S493 and S530,

may play important roles in directing assembly of the kinetochore. To discover proximal phosphorylation sites that may be part of kinase driven signalling pathways at the kinetochore, we also applied proximity phosphoproteomics in the presence of a covalent inhibitor of the kinetochore kinases CLK1 (KKT10)/ CLK2 (KKT19). We were able to detect a small subset of proximal phosphorylation sites that decreased at G2/M after KKT10/KKT19 inhibition with AB1 (Saldivia et al., 2020), including 3 affected phosphosites on KKT2 which is a known substrate of these kinases. Thus proximity phosphoproteomics enabled us to not only quantify phosphorylation sites at the kinetochore during the cell cycle, but also to observe the effect of kinase inhibition on phosphorylation sites within this complex, providing important information to begin understanding individual kinase mediated signalling pathways.

In our experiments we used a spatial reference to assess the background, this greatly assisted in determining which proteins and phosphorylation sites were proximal to the kinetochore kinases. We chose a chromatin reader, BDF5, which we endogenously tagged with BirA* or miniTurbo. BDF5 was an ideal spatial reference for the kinetochore, as although both are chromatin associated, they localise to distinct chromatin domains, allowing accurate determination of off-target biotinylation. BDF5 was also a protein of interest in its own right, thus BDF5 and the kinetochore acted as reciprocal spatial references allowing us to extract maximum biological insight from mass spectrometry time.

BDF5 is a predicted acetyl-lysine histone reader which accumulates at transcription start sites and plays an essential role in regulating transcription in *Leishmania* (Jones et al., 2021). Our XL-BiolD investigation revealed a core set of proximal proteins that included other bromodomain proteins BDF3, BDF4, a histone acetyltransferase HAT2, another predicted acetyl-lysine reader and several novel proteins. Evidence of a range of other chromatin/transcription associated processes occurring nearby was also captured by proximity biotinylation, such as detection of DNA replication and repair proteins, splicing factors, polyadenylation factors and RNA quality control proteins. Further analysis identified BDF5 as a regulator of transcription while proximal processes such as mRNA trans-splicing were unaffected. As with the kinetochore, we also discovered proximal phosphorylation sites which increased in number as parasites progressed through the cell cycle, although the overall number of phosphorylation sites detected was lower compared to the kinetochore.

Generating reliable spatial information: experimental design

In a proximity biotinylation experiment, direct interactors but also proximal proteins to the bait will become labelled. This is a distinct advantage of the method since it provides an insight into the molecular landscape surrounding a protein of interest. However, it means that a typical experiment may result in thousands of proteins being identified from the enriched biotinylated material. Determining which of these are proximal with high confidence will then depend on whether the experiment has been designed appropriately and also robust quantitative analysis of the mass spectrometry data. The overall aim when planning a proximity biotinylation experiment is to maximise the signal from proximal proteins whilst minimising background noise. Based on the examples in the literature we suggest careful consideration of the following points to increase the probability of a successful experiment:

1. Choice of enzyme. With an ever expanding repertoire of biotinylators to choose from, selecting the most suitable enzyme for the biological question may not be straightforward. A first consideration which might help narrow down the choice is the time scale that is being probed. If dynamic signalling networks are being investigated where a time resolution of minutes is required then an enzyme with fast kinetics such as APEX2 would be appropriate. If the goal is to characterize the steady state composition of a protein complex of interest, then BirA* with slower labelling kinetics may provide cleaner data for reasons outlined above. For intermediate cases, TurboID may be a useful enzyme, although the significant amount of

biotinylation that can take place even before adding biotin would increase background noise. The related miniTurbo enzyme may be considered with caution, as it appears to be less stable in some contexts (May et al., 2020). The concentration of exogenous biotin should be tailored to the chosen biotin ligase, up to 500µM biotin is used for miniTurbo or TurboID and 50-150µM biotin for BirA* ([Branon et al. 2018](#); [Geoghegan et al. 2021](#); [Zhou et al. 2018](#)).

2. Biotinylator location. Having selected an enzyme, the next decision is to determine where to place the biotinylator in the primary sequence of the bait protein. Factors to consider are common to the construction of any fusion protein. For N-terminal fusion, the sequence should be checked for the presence of any targeting signal that directs the proteins to a particular organelle such as the ER, since fusion to a biotinylator would likely disrupt targeting. Both termini should also be checked for domains known to be important for bait function. Ultimately though, the biotinylator-bait fusion protein should be expressed and correct localisation as well as function of the bait checked experimentally. Evidence that the biotinylator is active can be checked by streptavidin pull down from cultures supplemented or not with biotin, followed by western blotting to detect auto-biotinylation of the bait. Increased bait enrichment from biotin supplemented cultures, after normalising for input levels, indicates the biotinylator is active. To maximise coverage of the protein complex, it may be necessary to fuse the biotinylator to multiple components and possibly even using both N and C terminal fusions. This may also be helpful in identifying high confidence proximal proteins by analysing those proteins present in all datasets.
3. Expression level. Levels of the fusion protein should be as close to those of the endogenous protein as possible which may be achieved by endogenous tagging. It may be tempting to increase the signal to noise ratio by overexpressing the fusion protein, however this can lead to protein mis-localisation, generating a large amount of background biotinylated material which can obscure the true proximal proteins of interest and actually reduce signal to noise, as observed in a proximity biotinylation study of clathrin adaptor protein AP-1 comparing endogenous tagging with a biotinylator vs. transient overexpression (Stockhammer et al., 2021).
4. Spatial reference. This is probably the most important component of a proximity biotinylation experiment. The spatial reference consists of the biotinylating enzyme targeted to the same sub-cellular compartment as the biotinylator-bait fusion protein, providing a compartment specific background of incidentally biotinylated proteins against which enriched proteins from the bait sample can be compared (Lobingier et al., 2017). For example if the bait protein resides in the nucleus, the spatial reference could be the expression of the biotinylator containing a nuclear localisation signal. A potentially more efficient experimental design would be to choose another protein of interest to which the biotinylator is fused (**Fig. 2C**). Ideally this second protein would reside in the same compartment and at similar levels to the first bait, whilst being in a distinct complex. The two biotinylator-protein fusions could then act as reciprocal controls, making maximal use of experimental resources such as mass spectrometry time (Geoghegan et al., 2021; Jones et al., 2021).

Where resources permit, pilot proximity biotinylation experiments are an invaluable investment, enabling the above principles to be tested on the particular system of study. This could simply take the form of testing various biotinylator-bait fusion constructs for biotinylation activity by western blotting extracts with labelled streptavidin (e.g. HRP). Ideally though, pilot samples from biotinylator-bait fusions are analysed by mass spectrometry which will provide more detailed information on the signal to noise characteristics of the experiment. The bait protein should be one of the top identified hits after endogenously biotinylated proteins. If there is already some information on expected proximal proteins, the list should be checked for recall of these. Identification of a large number of chaperone proteins and other factors involved in protein folding may indicate that the cell has difficulty expressing and correctly processing the fusion protein (Banks et al., 2018).

Generating reliable spatial information: data analysis

Determining the proximal proteins from often quite complex datasets requires quantitative analysis of the enriched proteins. This firstly requires a number of biological replicates so that any enrichment observed against the control sample can be assessed relative to the variability of the measurement. At a minimum, 3 biological replicates should be analysed for each sample group, a higher number of replicates increases statistical power and may increase the sensitivity for detecting more transient proximal proteins. Semi-quantitative approaches to analysing proximity biotinylation data based on spectral counting have been successfully used (Lambert et al., 2015), but for optimal sensitivity a fully quantitative strategy should be considered which could be metabolic labelling (e.g. SILAC), chemical labelling (e.g. TMT) or label free quantitation.

In principle, detecting whether an identified protein is proximal is simply a matter of comparing its intensity in the bait samples to the intensity in the control samples, if it is significantly higher in the bait samples, it indicates the protein is proximal to the bait. There are a range of statistical methods for assessing whether enrichment is significant. The simplest approach is to carry out a pairwise statistical test, such as a t-test, between the intensities measured for a protein in the control and bait samples. This is performed protein-by-protein; therefore on a typical proximity biotinylation dataset, the statistical test is applied thousands of times and multiple testing correction must be applied to the p-values. While valid, this approach considers each protein in isolation, not in context of the quantitative information present in the entire dataset and therefore might lack sensitivity. Methods developed to analyse AP-MS data such as COMPASS, MiST and SAINTq (Sowa et al., 2009; Jäger et al., 2011; Teo et al., 2016) can instead be applied to proximity biotinylation data. In particular, SAINTq has emerged as a powerful probability modelling based tool for scoring interaction data. It can be used to score small and large scale proximity biotinylation studies and has the advantage of providing a False Discovery Rate (FDR) estimate for proximal proteins (Geoghegan et al., 2021).

Comparing intensities between samples and determining significant differences is a problem familiar to the microarray field. Several powerful statistical methods have been developed to analyse microarray data which can be leveraged to sensitively detect proximal proteins in proximity biotinylation data. The most well known of these is limma, an R package originally developed for analysing microarray data and later adapted for RNA-seq data (Ritchie et al., 2015). Limma provides a solution to the problem of variance estimation when there are a low number of biological replicates. A typical proximity biotinylation experiment may consist of 3 biological replicates in a control group and 3 biological replicates in a bait group. If a simple t-test was performed on each measured protein to determine the significance of a difference between the two groups, the variance is estimated from this small number of biological replicates for each protein. The variance estimation will not be very accurate and after performing this test on thousands of proteins, some proteins will appear to be significantly different as they happen to have a low variability. Conversely, those proteins which happen to have higher variability would not be called as significantly different and so may be missed. The key innovation in limma is that the variance for each protein is estimated from the measured variability of *all* proteins in the experiment in a process called empirical Bayes. This results in shrinking of estimated variance for higher variability proteins and increasing estimated variance for lower variability proteins. The overall effect is to make determination of significant differences between samples less dependent on chance variability.

Proximity biotinylation data may be visualised by plotting enrichment against p-values in a volcano plot. However, we suggest a more effective visualisation is to plot enrichment and p-values radially, plotting only data for proteins enriched against the control (Geoghegan et al., 2021). The bait protein normally has the lowest p-value and is therefore the central point in the plot, proximal proteins are then radially arranged close to the bait. Compared to volcano plots, this provides a better spread of proximal proteins which are the data of interest.

Anticipated improvements and future applications

Parasitologists were amongst the first to recognise and apply the power of proximity biotinylation to explore protein complexes and cellular processes, but it is still a relatively new technique under very active development. New biotin ligases will continue to be developed, likely drawing from sequences in metagenomic databases and using directed evolution to engineer desired characteristics such as fast labelling, small size, rapid folding, stability and possibly even a controllable labelling radius (Ke et al., 2021). Many biological systems including kinetoplastids express abundant endogenously biotinylated proteins which are unfortunately co-purified with proximity biotinylated proteins but are not of interest. Therefore a useful development would be a non-biotin proximity labelling system which would include a corresponding high affinity binder to enable specific purification of proximal material. Non-biotin proximity labelling methodologies do exist but usage is restricted due to non-diffusible labelling (PUP-IT) or requirement for a chemical catalyst (MicroMap) (Liu et al., 2018; Geri et al., 2020).

Proteins that form discrete structures have been the focus of most proximity biotinylation studies in kinetoplastids, but moving forward there are many interesting targets that do not form stable long-lived complexes and may have more diffuse localisations within the parasite. For example, these proteins may be part of signalling pathways that report on the parasite microenvironment in the host and instruct the parasite to differentiate. Conditional proximity biotinylation could be applied to proteins that have multiple localisations in the parasite, using recently developed split biotin ligase systems such as split TurboID (Cho et al., 2020). The bait would be fused to one half of the biotin ligase and another protein in the complex of interest is fused to the complementing biotin ligase portion. Close association between the two proteins re-constitutes biotin ligase activity so that only proximal proteins at the complex of interest are biotinylated. In theory this improves spatial specificity, but may also reduce signal as split enzymes often possess lower activity, it also requires knowledge of another interacting partner.

Another refinement is functional proximity biotinylation, where proximity material is enriched for proteins having certain biochemical characteristics or functions (Qin et al., 2021b). We demonstrated this in our proximity phosphoproteomics work on the *Leishmania* kinetochore. But proximal proteins bearing other PTMs could be enriched, such as glycosylation, acetylation, ubiquitin or SUMO. This interrogates PTMs within a complex of interest and is therefore a spatially focussed view of those modifications likely to be functionally important, as compared to many large scale PTM proteomics studies which are averages across whole cells or tissues. Finally, no studies have yet used proximity biotinylation to explore non-protein proximal biomolecules in kinetoplastids. Transcriptional regulation is largely absent in kinetoplastids, instead RNA binding proteins are the key regulators of mRNA levels (De Pablos et al., 2016). Exploring how RNA binding proteins modulate levels of certain transcripts could be possible by using APEX2, which can biotinylate proximal RNAs that can then be enriched and sequenced (Fazal et al., 2019). Non-biotinylation based labelling systems such as Cap-seq and Halo-seq could also be applied to investigate proximal RNAs (Wang et al., 2019; Engel et al., 2021). A recently introduced proximity labelling method, MicroMap, uses carbene based labelling (Geri et al., 2020). Carbenes are extremely reactive with a half life in the nanosecond range and were demonstrated to provide very spatially refined proximity labelling on the cell plasma membrane, they also react with C-H bonds and so could potentially label a very wide range of biomolecules such as RNA, DNA and lipids.

The intersection between chemistry and protein engineering will continue to expand the range of proximity labelling methods, offering kinetoplastid researchers a rich array of techniques with which to drive discovery in these unique parasites.

Acknowledgements: This work was funded by the Wellcome Trust (200807/Z/16/Z) and MRC GCRF (MR/P027989/1).

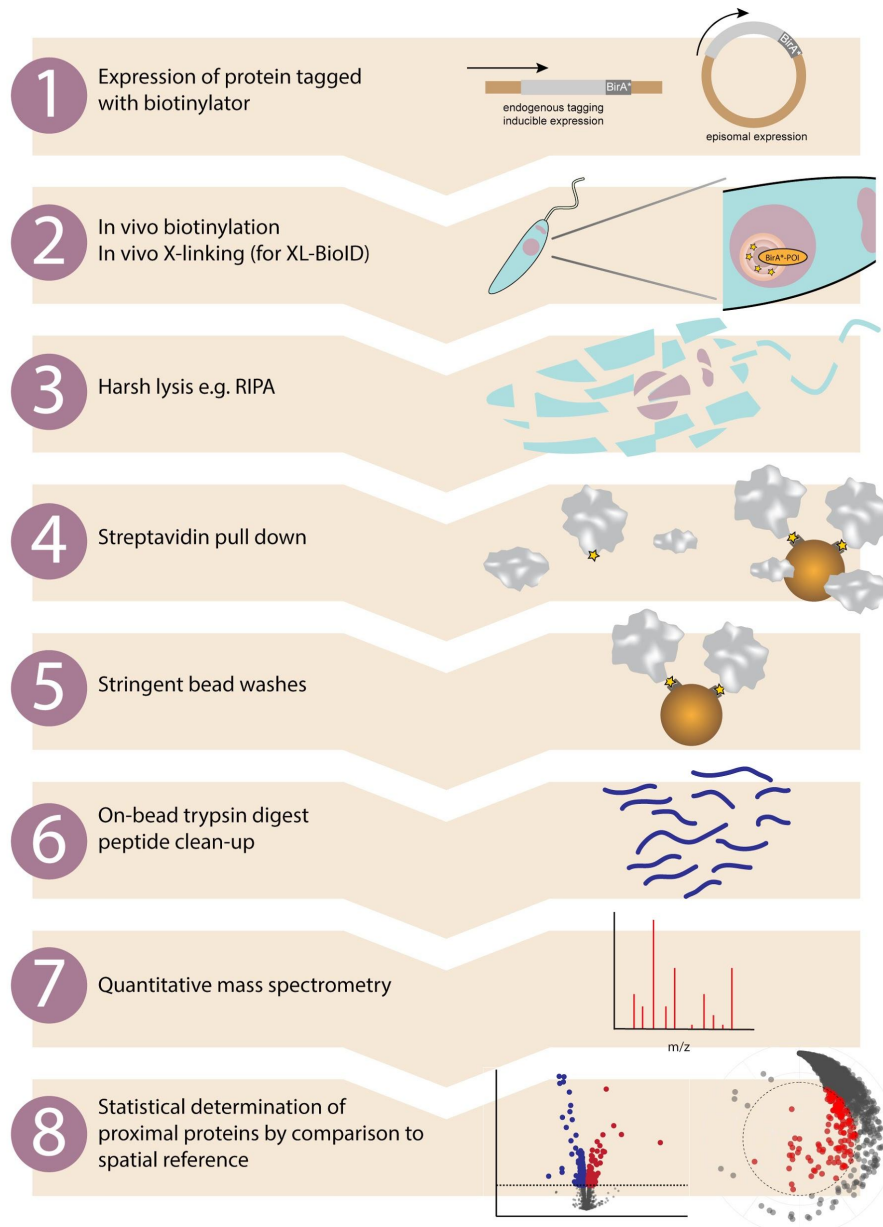


Figure 1: Proximity biotinylation workflow in kinetoplastids. Overview of the main experimental steps involved in a proximity biotinylation experiment. 1. Protein of interest is either tagged at the endogenous locus with a biotinylator such as BirA* or APEX2, or a tagged version is expressed from a different locus. 2. Adding exogenous biotin to the growth medium (BirA*) or biotin-phenol then H₂O₂ (APEX2) initiates *in vivo* proximity biotinylation. If performing XL-BioID, this may then be followed by an *in vivo* cross-linking step to increase capture of proximal proteins. A nuclear protein fused to BirA* is shown as an example. 3. The major advantage of proximity biotinylation is being able to completely lyse and solubilise cells with harsh lysis conditions such as radioimmunoprecipitation assay (RIPA) buffer. 4. Proximal proteins covalently tagged with biotin are enriched by incubating lysate with streptavidin beads. 5. Non-proximal proteins are washed off with a series of stringent bead washes. 6. Proteins remaining on beads are digested with trypsin and released peptides are desalted using C18. 7. Peptides are identified and quantified by quantitative mass spectrometry. Quantitation may be label free, TMT or SILAC based. 8. Enrichment of identified proteins vs. a control sample is calculated and statistical testing performed, e.g. with SAINTq or limma. Proteins significantly enriched in the sample vs. the spatial reference sample are classed as proximal proteins.

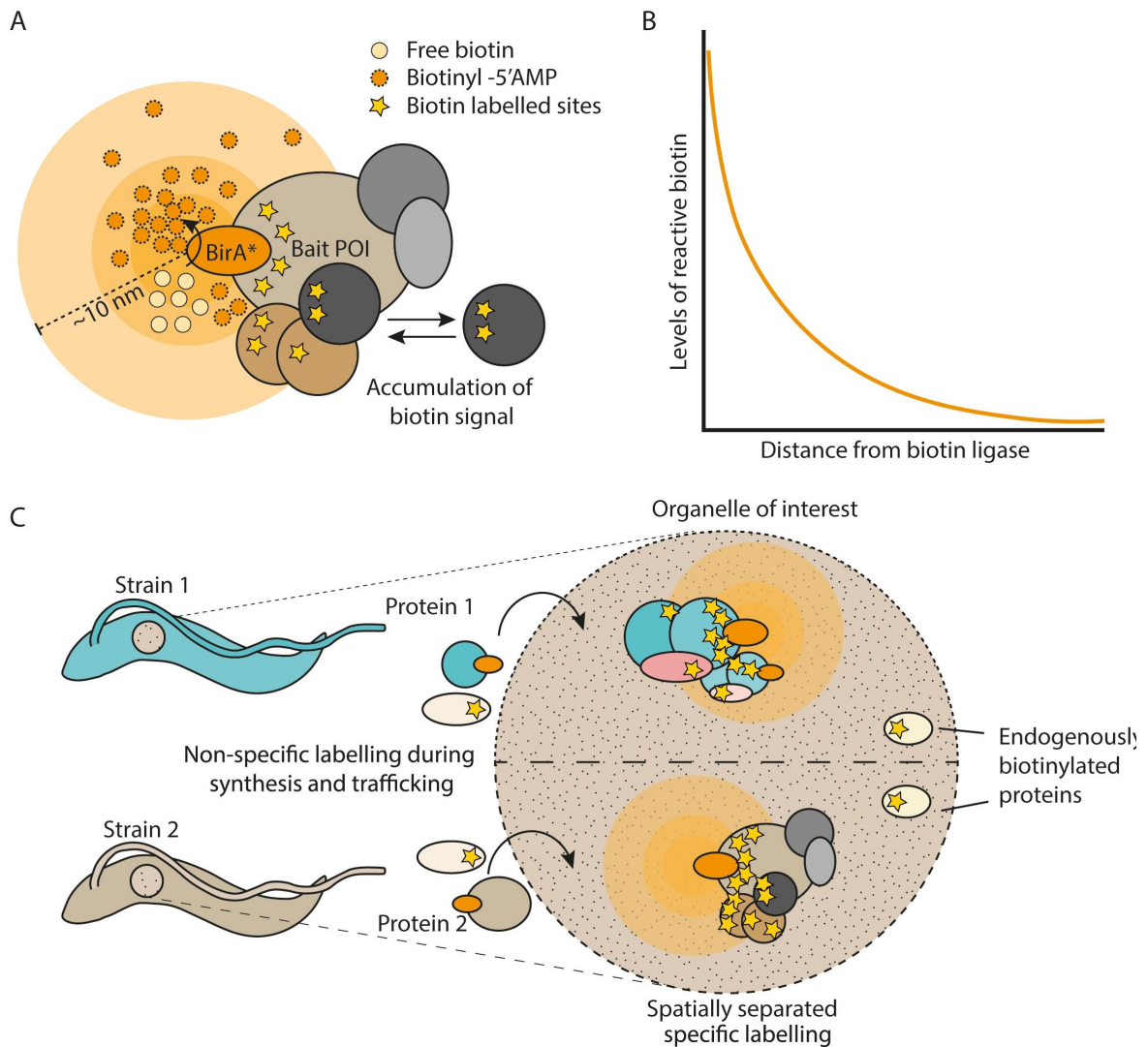
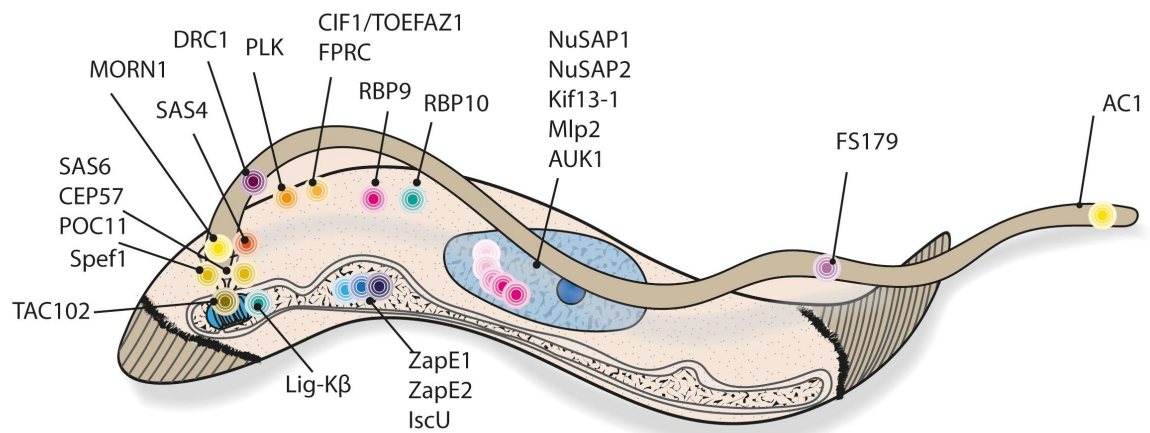


Figure 2: Overview of proximity biotinylation by BirA*. **A.** Cartoon depicting a bait protein of interest (POI) fused to BirA*. Free biotin is converted into a reactive biotinyl-5'AMP which reacts with lysine residues in proteins within a 10 nm labelling radius. Signal accumulates over time due to dynamic changes in the protein complex. **B.** Depiction of the exponential decay of reactive biotin levels as distance increases from the biotin ligase. **C.** Example of using spatial or organellar control strains, in which a different protein is tagged with the biotin ligase and compared against the strain expressing the control protein. This allows for the computational removal of endogenously biotinylated proteins as well as labelled proteins that are not specific to complexes of interest.

A



B

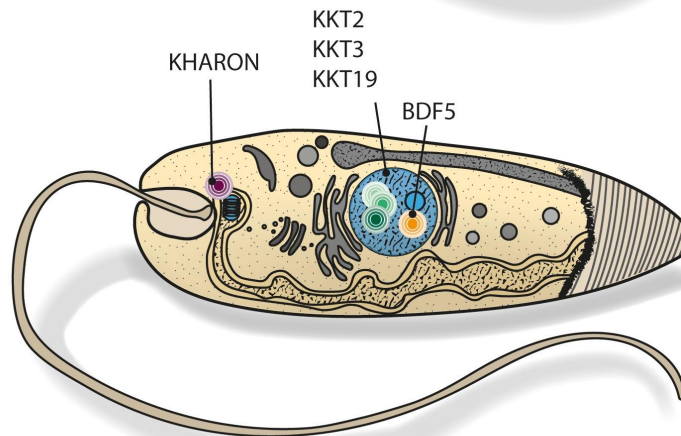


Figure 3: Cartoon depiction of kinetoplastid targets explored using proximity biotinylation. Cutaway style drawing, showing stylised cellular location of **A**, *T. brucei* and **B**, *Leishmania* proximity biotinylation bait proteins. N.B. many of the represented proteins have dynamic localisations throughout the cell cycle.

Table 1: List of proximity biotinylation studies conducted in kinetoplastids.

Bait Protein	Species	Gene ID	Bait Expression	Control	Enzyme/method	Reference
MORN1	<i>T. brucei</i> Procyclic Forms	<i>Tb927.6.4670</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Morriswood et al., 2013)
PLK	<i>T. brucei</i> Procyclic Forms	<i>Tb927.7.6310</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(McAllaster et al., 2015)
SAS-4	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.3300</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Hu et al., 2015)
CIF1	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.15800</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Zhou et al., 2016)
TbSAS-6, TbCEP57, TbPOC11	<i>T. brucei</i> Procyclic Forms	<i>Tb927.9.10550</i> , <i>Tb927.10.5990</i> , <i>Tb927.11.6560</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Dang et al., 2017)
RBP9, RBP10	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.12120</i> , <i>Tb927.8.2780</i>	Tet Inducible Allele	Strain Expressing GFP-BirA*	BirA*/BioID	(De Pablos et al., 2017)
NuSAP1, NuSAP2, Kif13-1, Mlp2, AUK1	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.8370</i> , <i>Tb927.9.6110</i> , <i>Tb927.9.3650</i> , <i>Tb927.9.1340</i> , <i>Tb927.11.8220</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Zhou et al., 2018b)
TOEFAZ1 (=CIF1)	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.15800</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Hilton et al., 2018)
CIF1	<i>T. brucei</i> Procyclic Forms	<i>Tb927.11.15800</i>	Tet Inducible Allele	Comparison to previous datasets	BirA*/BioID	(Zhou et al., 2018a)
FPRC	<i>T. brucei</i> Procyclic Forms	<i>Tb927.10.6360</i>	Tet Inducible Allele	Comparison to previous datasets	BirA*/BioID	(Hu et al., 2019)
ZapE1, ZapE2, IscU, LigK-beta	<i>T. brucei</i> Procyclic Forms	<i>Tb927.7.6930</i> , <i>Tb927.10.8070</i> , <i>Tb927.9.11720</i> , <i>Tb927.7.600</i>	Endogenous Tagging	Reciprocal organellar spatial controls	BioID2	(Pyrih et al., 2020)
FAZ27	<i>T. brucei</i> Procyclic Forms	<i>Tb927.9.8350</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(An et al., 2020)
Spef1	<i>T. brucei</i> Procyclic Forms	<i>Tb927.4.3130</i>	Endogenous Tagging	Wild Type Cells	BirA*/BioID	(Dong et al., 2020)
Unc119	<i>T. brucei</i> Procyclic Forms	<i>Tb927.2.4580</i>	Tet Inducible Alleles	Overlapping hits from N- and C-terminal tagged experiments	BioID2	(Pandey et al., 2020)
TAC102	<i>T. brucei</i> Procyclic Forms	<i>Tb927.7.2390</i>	Tet Inducible Allele	Plus/Minus Tet	BirA*/BioID	(Baudouin et al., 2020)
KHARON	<i>L. mexicana</i> Promastigotes	<i>LmxM.36.5850</i>	Episomal Allele	Wild Type Cells	BirA*/BioID	(Kelly et al., 2020)
DRC1, AC1, FS179, AC1Δ45	<i>T. brucei</i> Procyclic Forms	<i>Tb927.10.7880</i> , <i>Tb927.11.17040</i> , <i>Tb927.10.2880</i>	Endogenous Tagging	DRC1 vs WT, AC1/FS179 vs WT and AC1Δ45(cytosolic control)	APEX2	(Vélez-Ramírez et al., 2021)
KKT2, KKT3, KKT19	<i>L. mexicana</i> Promastigotes	<i>LmxM.36.5350</i> , <i>LmxM.34.4050</i> , <i>LmxM.09.0410</i>	Endogenous Tagging	Reciprocal organellar spatial controls	BirA*/miniTurbo/X L-BioID	(Geoghegan et al., 2021)
BDF5	<i>L. mexicana</i> Promastigotes	<i>LmxM.09.1260</i>	Endogenous Tagging	Reciprocal organellar spatial controls	BirA*/miniTurbo/X L-BioID	(Jones et al., 2021)

References

- Akiyoshi, B., and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156, 1247–1258.
- An, T., Zhou, Q., Hu, H., Cormaty, H., and Li, Z. (2020). FAZ27 cooperates with FLAM3 and ClpGM6 to maintain cell morphology in *Trypanosoma brucei*. *J. Cell Sci.* 133. doi:10.1242/jcs.245258.
- Banks, C. A. S., Miah, S., Adams, M. K., Eubanks, C. G., Thornton, J. L., Florens, L., et al. (2018). Differential HDAC1/2 network analysis reveals a role for prefoldin/CCT in HDAC1/2 complex assembly. *Sci. Rep.* 8, 13712.
- Bar, D. Z., Atkatsch, K., Tavarez, U., Erdos, M. R., Gruenbaum, Y., and Collins, F. S. (2018). Biotinylation by antibody recognition-a method for proximity labeling. *Nat. Methods* 15, 127–133.
- Baudouin, H. C. M., Pfeiffer, L., and Ochsenreiter, T. (2020). A comparison of three approaches for the discovery of novel tripartite attachment complex proteins in *Trypanosoma brucei*. *PLoS Negl. Trop. Dis.* 14, e0008568.
- Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., et al. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* 36, 880–887.
- Branon, T., Han, S., and Ting, A. (2017). Beyond Immunoprecipitation: Exploring New Interaction Spaces with Proximity Biotinylation. *Biochemistry* 56, 3297–3298.
- Burza, S., Croft, S. L., and Boelaert, M. (2018). Leishmaniasis. *Lancet* 392, 951–970.
- Choi-Rhee, E., Schulman, H., and Cronan, J. E. (2004). Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein Sci.* 13, 3043–3050.
- Cho, K. F., Branon, T. C., Rajeev, S., Svinkina, T., Udeshi, N. D., Thoudam, T., et al. (2020). Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. U. S. A.* 117, 12143–12154.
- Clayton, C. E. (2016). Gene expression in Kinetoplastids. *Curr. Opin. Microbiol.* 32, 46–51.
- Dang, H. Q., Zhou, Q., Rowlett, V. W., Hu, H., Lee, K. J., Margolin, W., et al. (2017). Proximity Interactions among Basal Body Components in *Trypanosoma brucei* Identify Novel Regulators of Basal Body Biogenesis and Inheritance. *MBio* 8. doi:10.1128/mBio.02120-16.
- De Pablos, L. M., Ferreira, T. R., and Walrad, P. B. (2016). Developmental differentiation in *Leishmania* lifecycle progression: post-transcriptional control conducts the orchestra. *Curr. Opin. Microbiol.* 34, 82–89.
- De Pablos, L. M., Kelly, S., de Freitas Nascimento, J., Sunter, J., and Carrington, M. (2017). Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding proteins in *Trypanosoma brucei*. *Open Biol.* 7. doi:10.1098/rsob.160159.
- Dong, X., Lim, T. K., Lin, Q., and He, C. Y. (2020). Basal Body Protein TbSAF1 Is Required for Microtubule Quartet Anchorage to the Basal Bodies in *Trypanosoma brucei*. *MBio* 11. doi:10.1128/mBio.00668-20.
- Engel, K. L., Lo, H.-Y. G., Goering, R., Li, Y., Spitale, R. C., and Taliaferro, J. M. (2021). Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *Nucleic Acids Res.* doi:10.1093/nar/gkab1185.
- Erickson, H. P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online* 11, 32–51.
- Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., et al. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* 178, 473–490.e26.

- Geoghegan, V., Jones, N. G., Dowle, A., and Mottram, J. C. (2021). Protein kinase signalling at the Leishmania kinetochore captured by XL-BioID. *bioRxiv*, 2021.07.08.451598. doi:10.1101/2021.07.08.451598.
- Geri, J. B., Oakley, J. V., Reyes-Robles, T., Wang, T., McCarver, S. J., White, C. H., et al. (2020). Microenvironment mapping via Dexter energy transfer on immune cells. *Science* 367, 1091–1097.
- Gingras, A.-C., Abe, K. T., and Raught, B. (2019). Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr. Opin. Chem. Biol.* 48, 44–54.
- Gingras, A.-C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645–654.
- Hilton, N. A., Sladewski, T. E., Perry, J. A., Pataki, Z., Sinclair-Davis, A. N., Muniz, R. S., et al. (2018). Identification of TOEFAZ1-interacting proteins reveals key regulators of Trypanosoma brucei cytokinesis. *Mol. Microbiol.* 109, 306–326.
- Hopkins, C., Gibson, A., Stinchcombe, J., and Futter, C. (2000). Chimeric molecules employing horseradish peroxidase as reporter enzyme for protein localization in the electron microscope. *Methods Enzymol.* 327, 35–45.
- Hu, H., An, T., Kurasawa, Y., Zhou, Q., and Li, Z. (2019). The trypanosome-specific proteins FPRC and CIF4 regulate cytokinesis initiation by recruiting CIF1 to the cytokinesis initiation site. *J. Biol. Chem.* 294, 16672–16683.
- Hu, H., Zhou, Q., and Li, Z. (2015). SAS-4 Protein in Trypanosoma brucei Controls Life Cycle Transitions by Modulating the Length of the Flagellum Attachment Zone Filament. *J. Biol. Chem.* 290, 30453–30463.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., et al. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 4, 1265–1272.
- Jäger, S., Cimerancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., et al. (2011). Global landscape of HIV–human protein complexes. *Nature* 481, 365–370.
- Jones, N. G., Geoghegan, V., Moore, G., Carnielli, J. B. T., Newling, K., Calderón, F., et al. (2021). Bromodomain factor 5 is an essential transcriptional regulator of the Leishmania genome. *bioRxiv*, 2021.09.29.462384. doi:10.1101/2021.09.29.462384.
- Kelly, F. D., Tran, K. D., Hatfield, J., Schmidt, K., Sanchez, M. A., and Landfear, S. M. (2020). A cytoskeletal protein complex is essential for division of intracellular amastigotes of Leishmania mexicana. *J. Biol. Chem.* 295, 13106–13122.
- Ke, M., Yuan, X., He, A., Yu, P., Chen, W., Shi, Y., et al. (2021). Spatiotemporal profiling of cytosolic signaling complexes in living cells by selective proximity proteomics. *Nat. Commun.* 12, 71.
- Kido, K., Yamanaka, S., Nakano, S., Motani, K., Shinohara, S., Nozawa, A., et al. (2020). AirID, a novel proximity biotinylation enzyme, for analysis of protein–protein interactions. *Elife* 9, e54983.
- Kim, D. I., Birendra, K. C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2453–61.
- Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., et al. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* 27, 1188–1196.
- Kostygov, A. Y., Karnkowska, A., Votýpka, J., Tashyreva, D., Maciszewski, K., Yurchenko, V., et al. (2021). Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 11,

200407.

- Kwon, K., and Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci.* 9, 1530–1539.
- Lambert, J.-P., Tucholska, M., Go, C., Knight, J. D. R., and Gingras, A.-C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics* 118, 81–94.
- Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., et al. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* 12, 51–54.
- Libisch, M. G., Rego, N., and Robello, C. (2021). Transcriptional Studies on *Trypanosoma cruzi* – Host Cell Interactions: A Complex Puzzle of Variables. *Frontiers in Cellular and Infection Microbiology* 11. doi:10.3389/fcimb.2021.692134.
- Liu, Q., Zheng, J., Sun, W., Huo, Y., Zhang, L., Hao, P., et al. (2018). A proximity-tagging system to identify membrane protein-protein interactions. *Nat. Methods* 15, 715–722.
- Llauró, A., Hayashi, H., Bailey, M. E., Wilson, A., Ludzia, P., Asbury, C. L., et al. (2018). The kinetoplastid kinetochore protein KKT4 is an unconventional microtubule tip-coupling protein. *J. Cell Biol.* 217, 3886–3900.
- Lobingier, B. T., Hüttenhain, R., Eichel, K., Miller, K. B., Ting, A. Y., von Zastrow, M., et al. (2017). An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* 169, 350–360.e12.
- May, D. G., Scott, K. L., Campos, A. R., and Roux, K. J. (2020). Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. *Cells* 9. doi:10.3390/cells9051070.
- McAllaster, M. R., Ikeda, K. N., Lozano-Núñez, A., Anrather, D., Unterwurzacher, V., Gossenreiter, T., et al. (2015). Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in *Trypanosoma brucei*. *Mol. Biol. Cell* 26, 3013–3029.
- Mehla, J., Caufield, J. H., Sakhawalkar, N., and Uetz, P. (2017). A comparison of two-hybrid approaches for detecting protein-protein interactions. *Methods Enzymol.* 586, 333–358.
- Miyagawa-Yamaguchi, A., Kotani, N., and Honke, K. (2014). Segregation of Lipid Rafts Revealed by the EMARS Method Using GPI-Anchored HRP Fusion Proteins. *Trends in Glycoscience and Glycotechnology* 26, 59–69. doi:10.4052/tigg.26.59.
- Morriswood, B., Havlicek, K., Demmel, L., Yavuz, S., Sealey-Cardona, M., Vidilaseris, K., et al. (2013). Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryot. Cell* 12, 356–367.
- Paek, J., Kalocsay, M., Staus, D. P., Wingler, L., Pascolutti, R., Paulo, J. A., et al. (2017). Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* 169, 338–349.e11.
- Pandey, M., Huang, Y., Lim, T. K., Lin, Q., and He, C. Y. (2020). Flagellar targeting of an arginine kinase requires a conserved lipidated protein intraflagellar transport (LIFT) pathway in *Trypanosoma brucei*. *J. Biol. Chem.* 295, 11326–11336.
- Pays, E., and Nolan, D. P. (2021). Genetic and immunological basis of human African trypanosomiasis. *Curr. Opin. Immunol.* 72, 13–20.
- Povelones, M. L. (2014). Beyond replication: division and segregation of mitochondrial DNA in kinetoplastids. *Mol. Biochem. Parasitol.* 196, 53–60.
- Pyrih, J., Rašková, V., Škodová-Sveráková, I., Pánek, T., and Lukeš, J. (2020). ZapE/Afg1 interacts with Oxa1 and its depletion causes a multifaceted phenotype. *PLoS One* 15, e0234918.

- Qin, W., Cho, K. F., Cavanagh, P. E., and Ting, A. Y. (2021a). Deciphering molecular interactions by proximity labeling. *Nat. Methods* 18, 133–143.
- Qin, W., Myers, S. A., Carey, D. K., Carr, S. A., and Ting, A. Y. (2021b). Spatiotemporally-resolved mapping of RNA binding proteins via functional proximity labeling reveals a mitochondrial mRNA anchor promoting stress recovery. *Nat. Commun.* 12, 4980.
- Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., et al. (2018). RNA-protein interaction detection in living cells. *Nat. Methods* 15, 207–212.
- Remnant, L., Booth, D. G., Vargiu, G., Spanos, C., Kerr, A. R. W., and Earnshaw, W. C. (2019). In vitro BioID: mapping the CENP-A microenvironment with high temporal and spatial resolution. *Mol. Biol. Cell* 30, 1314–1325.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Roux, K. J., Kim, D. I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* 196, 801–810.
- Saldivia, M., Fang, E., Ma, X., Myburgh, E., Carnielli, J. B. T., Bower-Lepts, C., et al. (2020). Targeting the trypanosome kinetochore with CLK1 protein kinase inhibitors. *Nat Microbiol* 5, 1207–1216.
- Samavarchi-Tehrani, P., Samson, R., and Gingras, A.-C. (2020). Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Mol. Cell. Proteomics* 19, 757–773.
- Sowa, M. E., Bennett, E. J., Gygi, S. P., and Harper, J. W. (2009). Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138, 389–403.
- Stockhammer, A., Benz, L. S., Freund, C., Kuropka, B., and Bottanelli, F. (2021). When less is more - Endogenous tagging with TurboID increases the sensitivity of proximity labelling-based experiments. *bioRxiv*, 2021.11.19.469212. doi:10.1101/2021.11.19.469212.
- Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., et al. (2006). A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol. Cell. Proteomics* 5, 737–748.
- Teo, G., Koh, H., Fermin, D., Lambert, J.-P., Knight, J. D. R., Gingras, A.-C., et al. (2016). SAINTq: Scoring protein-protein interactions in affinity purification - mass spectrometry experiments with fragment or peptide intensity data. *Proteomics* 16, 2238–2245.
- Tiengwe, C., Marcello, L., Farr, H., Dickens, N., Kelly, S., Swiderski, M., et al. (2012). Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep.* 2, 185–197.
- Vélez-Ramírez, D. E., Shimogawa, M. M., Ray, S. S., Lopez, A., Rayatpisheh, S., Langousis, G., et al. (2021). APEX2 Proximity Proteomics Resolves Flagellum Subdomains and Identifies Flagellum Tip-Specific Proteins in *Trypanosoma brucei*. *mSphere* 6. doi:10.1128/mSphere.01090-20.
- Wang, P., Tang, W., Li, Z., Zou, Z., Zhou, Y., Li, R., et al. (2019). Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat. Chem. Biol.* 15, 1110–1119.
- Wartlick, O., Kicheva, A., and González-Gaitán, M. (2009). Morphogen gradient formation. *Cold Spring Harb. Perspect. Biol.* 1, a001255.
- Zhao, X., Bitsch, S., Kubitz, L., Schmitt, K., Deweid, L., Roehrig, A., et al. (2021). ultraID: a compact and efficient enzyme for proximity-dependent biotinylation in living cells. *bioRxiv*, 2021.06.16.448656. doi:10.1101/2021.06.16.448656.
- Zhou, Q., An, T., Pham, K. T. M., Hu, H., and Li, Z. (2018a). The CIF1 protein is a master orchestrator of trypanosome cytokinesis that recruits several cytokinesis regulators to the cytokinesis initiation site. *J. Biol. Chem.* 293, 16177–16192.

- Zhou, Q., Hu, H., and Li, Z. (2016). An EF-hand-containing Protein in *Trypanosoma brucei* Regulates Cytokinesis Initiation by Maintaining the Stability of the Cytokinesis Initiation Factor CIF1. *J. Biol. Chem.* 291, 14395–14409.
- Zhou, Q., Lee, K. J., Kurasawa, Y., Hu, H., An, T., and Li, Z. (2018b). Faithful chromosome segregation in *Trypanosoma brucei* requires a cohort of divergent spindle-associated proteins with distinct functions. *Nucleic Acids Res.* 46, 8216–8231.