Published in final edited form as:

J Proteomics. 2015 April 6; 118: 81–94. doi:10.1016/j.jprot.2014.09.011.

Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes

Jean-Philippe Lambert¹, Monika Tucholska¹, Christopher Go^{1,2}, James D.R. Knight¹, and Anne-Claude Gingras^{1,2}

¹Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, M5G 1X5, Canada

²Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada

Abstract

Mapping protein-protein interactions for chromatin-associated proteins remains challenging. Here we explore the use of BioID, a proximity biotinylation approach in which a mutated biotin ligase (BirA*) is fused to a bait of interest, allowing for the local activation of biotin and subsequent biotinylation of proteins in the bait vicinity. BioID allowed for successful interactome mapping of core histones and members of the mediator complex. We explored the background signal produced by the BioID approach and found that using distinct types of controls increased the stringency of our statistical analysis with *SAINTexpress*. A direct comparison of BioID with our AP-MS protocol optimized for chromatin-associated protein complexes revealed that the approaches identified few shared interaction partners and enriched for distinct biological processes; yet, both approaches permitted the recovery of biologically meaningful interactions. While no clear bias could be observed for either technique toward protein complexes of particular functions, BioID allowed for the purification of proteins of lower cellular abundance. Finally, we were able to identify a strong association of MED4 with the centrosome by BioID and validated this finding by immunofluorescence. In summary, BioID complements AP-MS for the study of chromatin-associated protein complexes.

Keywords

Proximity biotinylation; BioID; systems biology; chromatin; protein-protein interactions; affinity purification coupled to mass spectrometry; mediator complex

© 2014 Elsevier B.V. All rights reserved.

Please send correspondence to: Anne-Claude Gingras, gingras@lunenfeld.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Chromatin, the protein and DNA fractions composing chromosomes, is central to the preservation and utilization of a cell's genetic material. Numerous proteins are known to assemble into large complexes that participate in proper chromatin structure (e.g. histone chaperones and nucleosome remodelers) and efficient gene transcription (e.g. transcription factors and epigenetic readers). Consequently, numerous approaches have been developed to identify protein complexes associated with chromatin, reviewed in [1]. Affinity purification coupled to mass spectrometry (AP-MS) is particularly well suited for these analyses. AP-MS is often performed by expressing the protein of interest, termed the bait, fused to an epitope tag that facilitates affinity purification, although this approach is also amenable to the analysis of endogenous baits if appropriate antibodies are available [2]. Following cell lysis, antibodies conjugated to a solid support are used to isolate the soluble bait (and associated preys) from the cell extract, enabling subsequent analysis by mass spectrometry.

A key determinant of AP-MS success is the solubility, or availability, of the protein complex under study. While this is not generally an obstacle for cytosolic protein complexes, it can be problematic when studying chromatin-associated protein complexes [3]. To circumvent this issue, two main strategies have emerged, namely salt extraction and chromatin shearing. By increasing the concentration of sodium or potassium chloride in the lysis buffer to 420 mM, it is possible to dissociate histone proteins from DNA and thus drastically increase the solubility of chromatin associated protein complexes [4]. This method was successfully applied in the characterization of 293 baits, resulting in the identification of 164 protein complexes [5]. A major downside of using high salt concentrations, however, is that many protein-protein interactions are not stable under these conditions, resulting in a high rate of false negatives. Chromatin shearing, on the other hand, can be achieved by using nucleases, sonication or both [1], with the advantage that salt-sensitive interactions should be better preserved. This can be exploited to identify direct interaction partners and also protein complexes linked to the bait by short stretches of DNA [6, 7]. Beyond the solubility of a protein complex, another major determinant of successful AP-MS is the rate of dissociation of the protein complex under study throughout the purification procedure [8]. Interaction partners that are readily lost through purification can be stabilized by chemically crosslinking them [9]. This approach has been successfully used in a large-scale study [10], but is generally avoided due to the optimization required for each bait.

BioID, an alternative approach to AP-MS that relies on proximity-dependent protein biotinylation, was recently introduced by Roux *et al.* [11]. In BioID, an *E. coli* biotin protein ligase harboring an R118G mutation, referred herein as BirA*, is fused in frame to a protein of interest. The R118G BirA* mutant can still catalyze the formation of activated biotin (biotinoyl-5'-AMP) but quickly dissociates from this intermediate [12]. The BirA* tagged bait therefore generates a cloud of activated biotin *in vivo*, which in turn can react with free primary amines, notably with the epsilon amine of lysine residues (Figure 1). As interaction partners and neighbors are marked by stable covalent modifications of their lysine side-chains, it is unnecessary to maintain protein complexes throughout the purification scheme. Harsh lysis conditions can thus be employed to effectively solubilize most cellular proteins. Subsequently, interaction partners of a BirA*-tagged bait can be effectively enriched for by

performing a purification with streptavidin (which avidly binds to biotin) coupled to mass spectrometric analysis. To date, the BioID approach has been applied to baits located at the nuclear lamina [11], the centrosome [13, 14], cytoskeleton [15], nuclear pore [16] and recently by our group to study proteins within the Hippo pathway [17]. We observed BioID enabled the detection of known interaction partners of the transcriptional co-activator YAP1, for example the TEAD family transcription factors, but also novel protein-protein interactions with SWI/SNF complex subunits and other chromatin-associated proteins [17].

This success led us to investigate how generally applicable the BioID approach is to chromatin-associated proteins. Here we further explore the background associated with BioID using three distinct types of controls. In addition, we performed a direct comparison for two distinct groups of chromatin-associated proteins, namely two histone proteins and three mediator complex subunits, using both AP-MS and BioID (Figure 1). Analysis of histone H3 by BioID and AP-MS produced largely different interactome maps enriching for distinct sets of biological processes. Further analysis also revealed that BioID had less of an abundance bias than AP-MS for the recovery of histone-associated proteins. The detailed study of three mediator complex subunits showed that selecting a given subunit within a large protein complex for tagging influenced the results of interactome mapping. Intriguingly, we also discovered that at least one subunit of the mediator complex, MED4, localizes to the centrosome by BioID, AP-MS and immunofluorescence. Overall, we observed that interactions identified with BioID displayed a rather small overlap with AP-MS data, but enabled access to a novel interactome space that is consistent with the function of these chromatin-associated proteins.

Methods

Construct and stable cell line generation

Constructs for the genes of interest were generated via Gateway cloning into pDEST 5' Triple-FLAG-pcDNA5-FRT-TO or pDEST 5' BirA*-FLAG-pcDNA5-FRT-TO. Entry clones for HIST1H2BG (accession # EU446968), HIST1H3A (accession # HQ448409), MED4 (accession # DQ893076), MED20 (accession # BC012618) and MED23 (accession # EU832308) were obtained from the ORFeome collection (<http://horfdb.dfci.harvard.edu/>) [18, 19], archived at the Lunenfeld-Tanenbaum using OpenFreezer [20]. Bait proteins of interest were stably expressed in T-REx Flp-In HEK293 cells as described [17]. Parental Flp-In T-REx HEK293 cells, and stable cells expressing BirA*-FLAG fused either to a green fluorescent protein (GFP) or to a nuclear localization sequence (NLS) were used as negative controls for the BioID experiments and processed in parallel to the bait proteins. Parental Flp-In T-REx HEK293 cells and cells expressing NLS-BirA* fused to FLAG tag were used as negative controls for AP-MS experiments and processed in parallel to the bait-expressing cell lines. Stable cell lines were selectively grown in the presence of 200 µg/mL hygromycin up to 80% confluence before expression was induced via 1 µg/mL tetracycline for 24 hours and the cells were harvested. For the BioID experiments, 50 µM biotin was added at the time of induction. Two 150-mm plates were induced with tetracycline and treated with biotin for 24 hours before harvesting. Cells were pelleted at low speed, washed with ice-cold PBS and frozen at -80°C until purification.

Affinity purification coupled to mass spectrometry

The FLAG AP-MS protocol was adapted from [3] with slight modifications. Stable cells from two 150-mm plates were pelleted, frozen down and lysed in 1.5 mL ice cold 50 mM HEPES-NaOH pH 8.0, 100 mM KCl, 2 mM EDTA, 0.1% NP40, and 10% glycerol with 1 mM PMSF, 1 mM DTT and Sigma protease inhibitor cocktail (P8340, 1:500) added immediately prior to processing. To aid with lysis, the cells were frozen on dry ice and thawed in a 37°C water bath, and then put back on ice. The samples were sonicated at 4°C using three 10 sec bursts with 2 sec pauses at 35% amplitude. 100 units of benzonase was then added and the lysates were incubated at 4°C for an hour with rotation. The lysates were then centrifuged at 20,817 g for 20 min at 4°C and the supernatant was then added to tubes containing 25 µl of 50% magnetic anti-FLAG M2 bead (Sigma, M8823) slurry prewashed in lysis buffer. FLAG immunoprecipitation was allowed to proceed at 4°C for 2 hours with rotation. Beads were pelleted by centrifugation (1000 rpm for 1 min) and magnetized, then the unbound lysate was aspirated and kept for analysis. The beads were demagnetized and washed with 1 mL lysis buffer and magnetized to aspirate off the wash buffer. The beads were then washed with 1 mL of 20 mM Tris-HCl (pH 8.0) containing 2 mM CaCl₂, then any excess wash buffer was removed by centrifuging the beads, magnetizing and pipetting off the liquid. The now-dry magnetic beads were removed from the magnet and resuspended in 7.5 µl of 20 mM Tris-HCl (pH 8.0) containing 750 ng of trypsin (Sigma, T7575) and the mixture was incubated at 37°C with agitation overnight. After the initial incubation, the beads were magnetized and the supernatant was transferred to a fresh tube. Another 250 ng of trypsin was added to the mixture and further digested, without agitation, for 3–4 hours. The sample was acidified with formic acid to a final concentration of 2% and the tryptic digests were stored at –40°C until ready for mass spectrometry analysis.

Proximity biotinylation coupled to mass spectrometry

Cell pellets were thawed in 1.5 mL ice cold RIPA buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0.1% SDS and 0.5% sodium deoxycholate. PMSF (1 mM), DTT (1 mM) and Sigma protease inhibitor cocktail (P8340, 1:500) were added immediately before use. The lysates were sonicated, treated with benzonase and centrifuged as described in the FLAG AP-MS section. For each sample, 60 µL of streptavidin-sepharose bead slurry (GE Healthcare, Cat 17-5113-01) were pre-washed three times with 1 mL of lysis buffer by pelleting the beads with gentle centrifugation and aspirating off the supernatant before adding the next wash. Biotinylated proteins were captured on pre-washed streptavidin beads for 3 hours at 4°C with rotation. The beads were then gently pelleted and the unbound supernatant was saved for further analysis. The beads were then washed 2 × 1 mL with RIPA buffer and 3 × 1 mL with 50 mM ammonium bicarbonate (pH 8.0). Following the final wash, the beads were pelleted and any excess liquid was aspirated off. Beads were then resuspended in 100 µL of 50 mM ammonium bicarbonate, and 1 µg of trypsin solution was added. The samples were incubated overnight at 37°C with rotation and then an additional 1 µg of trypsin was added, followed by a further incubation for 2–4 hours. The beads were pelleted and the supernatant was transferred to a fresh tube. The beads were rinsed with 2 × 100 µL HPLC grade water and the wash fraction was combined with the supernatant. The peptide solution was acidified with 50% formic acid to a final concentration of 2% and the samples were placed in a speedvac to dry.

Tryptic peptides were resuspended in 25 μ L 5% formic acid and stored at -80°C until analyzed by mass spectrometry.

Mass spectrometry analysis

AP-MS samples, BioID samples and controls were analyzed by mass spectrometry in at least two biological replicates. 5 μ L of each sample was directly loaded at 400 nL/min onto a 75 $\mu\text{m} \times 12$ cm emitter packed with 3 μm ReproSil-Pur C₁₈-AQ (Dr. Maisch HPLC GmbH, Germany). The peptides were eluted from the column over a 90 min gradient generated by a NanoLC-Ultra 1D plus (Eksigent, Dublin CA) nano-pump and analyzed on a TripleTOFTM 5600 instrument (AB SCIEX, Concord, Ontario, Canada). The gradient was delivered at 200 nL/min starting from 2% acetonitrile with 0.1% formic acid to 35% acetonitrile with 0.1% formic acid over 90 minutes followed by a 15 min clean-up at 80% acetonitrile with 0.1% formic acid, and a 15 min equilibration period back to 2% acetonitrile with 0.1% formic acid for a total of 120 min. To minimize carryover between each sample, the analytical column was washed for 3 hours by running an alternating sawtooth gradient from 35% acetonitrile with 0.1% formic acid to 80% acetonitrile with 0.1% formic acid, holding each gradient concentration for 5 min. Analytical column and instrument performance were verified after each sample by loading 30 fmol BSA tryptic peptide standard (Michrom Bioresources Inc. Fremont, CA) with 60 fmol α -Casein tryptic digest and running a short 30 min gradient. TOF MS calibration was performed on BSA reference ions before running the next sample in order to adjust for mass drift and verify peak intensity. The instrument method was set to a discovery or IDA mode which consisted of one 250 ms MS1 TOF survey scan from 400–1300 Da followed by twenty 100 ms MS2 candidate ion scans from 100–2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 4+ which exceeded a threshold of 200 cps were selected for MS2, and former precursors were excluded for 10 secs after 1 occurrence.

MS data analysis

Mass spectrometry data generated by TripleTOFTM 5600 were stored, searched and analyzed using the ProHits laboratory information management system (LIMS) platform [21]. Within ProHits, the resulting WIFF files were first converted to an MGF format using WIFF2MGF converter and to an mzML format using ProteoWizard (v3.0.4468) and the AB SCIEX MS Data Converter (V1.3 beta) and then searched using Mascot (v2.3.02) and Comet (v2012.02 rev.0). The spectra were searched with the RefSeq database (version 53, May 28th, 2014) acquired from NCBI against a total of 34374 human and adenovirus sequences supplemented with “common contaminants” from the Max Planck Institute (<http://maxquant.org/downloads.htm>) and the Global Proteome Machine (GPM; <http://www.thegpm.org/crap/index.html>). The database parameters were set to search for tryptic cleavages, allowing up to 2 missed cleavage sites per peptide with a mass tolerance of 40 ppm for precursors with charges of 2+ to 4+ and a tolerance of ± 0.15 amu for fragment ions. Variable modifications were selected for deamidated asparagine and glutamine and oxidized methionine. The results from each search engine were analyzed through TPP (the Trans-Proteomic Pipeline, [22] (v4.6 OCCUPY rev 3) via the iProphet pipeline [23]. SAINTexpress version 3.3 [24] was used as a statistical tool to calculate the probability value of each potential protein-protein interaction from background contaminants using default parameters. Unless otherwise specified, controls were compressed to 3 samples and

two unique peptides ions and a minimum iProphet probability of 0.95 were required for protein identification.

Functional enrichment and data visualization

Functional enrichment analysis was performed using DAVID Bioinformatics Resources 6.7 [25]. We performed Functional Annotation Clustering (June 2014) and reported Biological Process (BP FAT) enrichment probabilities adjusted for GO terms with Benjamini-Hochberg FDR correction of 0.01 or smaller. Heatmaps were generated with MultiExperiment Viewer version 4.8 (MeV; <http://www.tm4.org/mev/>) using hierarchical clustering with Pearson correlation and average linkage. Prey proteins with SAINT BFDR of 1% were used for heat map generation. Venn diagrams were generated using a local implementation of chart wizard (Google). Dot plots of histone H2B and H3 interactomes obtained by BioID and AP-MS were generated using the R statistical package (V 3.1.0) using custom-made scripts (Knight et al., submitted) [26]. Box plot showing the PaxDB abundance (integrated dataset) of prey proteins associated with H2B and H3 by BioID and AP-MS were generated using BoxPlotR [27].

Data deposition in public repositories

All control samples were deposited in the Contaminant Repository for Affinity Purification (www.crapome.org) [28]. The empty Flp-In T-REx HEK293 BioID controls are CC675, CC676, CC680, CC681, CC686 and CC687 and were annotated with protocol #160. The NLS-BirA*-FLAG BioID controls are CC668, CC672 and CC673 with annotated protocol #158 while the GFP-BirA*-FLAG BioID controls are CC674, CC677, CC678, CC679, CC682, CC683, CC684, CC685, CC688, CC689, CC690 and CC691 using protocol #161. The empty 3XFLAG AP-MS controls were previously deposited as CC564, CC568 and CC569 using protocol #142 while the NLS-BirA*-FLAG (no biotin added) AP-MS controls were deposited as CC669, CC670 and CC671 and annotated with protocol #159. All MS files used in this study were also deposited at MassIVE (<http://massive.ucsd.edu>). The MassIVE ID is MSV000078782 and the MassIVE link for download is <http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=3e4c7efd6da945d0a3fa7960f8a05df0>. The password for download prior to final acceptance is JPROT2014.

Western blot

Proteins were separated by SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were blocked in TBS containing 5 mg/mL non-fat milk and 1% Tween 20 for 1 hour at room temperature. Blots were probed for FLAG (1:5000; F1804; Sigma-Aldrich), β -tubulin (1:5000; E7; DSHB at the University of Iowa), H3K27ac (1:1000; 07-360; EMD Millipore) and Streptavidin conjugated to HRP (1:5000; RPN1231V; GE Healthcare). Detection on film was performed by chemiluminescence using the LumiGLO reagent (Cell Signaling Technology; #7003; 1:20).

Immunofluorescence

Flp-In T-REx HEK293 cells stably expressing 3XFLAG or BirA*-FLAG tagged proteins were seeded on poly-L-lysine coated cover slips (product #354085; BD Biosciences) at low

density and grown in complete medium for 24 hours. Cells were fixed with 3.7% paraformaldehyde/PBS and permeabilized in 0.3% Triton X-100 in PBS. Mouse anti-FLAG M2 antibody (1:2000; F1804, Sigma-Aldrich), rabbit anti-FLAG antibody (1:1000; F7425, Sigma-Aldrich), mouse anti-centrin (1:1000, 04-1624, EMD Millipore), rabbit anti-pericentrin (1:2000, ab4448, Abcam) and streptavidin Alexa-Fluor 488 (1:2500; S32354; Invitrogen) were used to identify FLAG tagged proteins and biotinylated proteins, respectively. Proteins were visualized with goat anti-mouse or anti-rabbit coupled to Alexa-Fluor 488 or 555 antibodies (1:1,000; A11001, A11008, A21422, A21428; Invitrogen). DNA was detected with DAPI staining. Immunofluorescence was observed by confocal microscopy on a Nikon Eclipse C1si instrument.

Results and Discussions

Exploring background proteins associated with BioID

While we and others have successfully used BioID for interaction mapping, a thorough investigation on the effects of protein expression levels and subcellular localization on the background proteins identified by the approach was still lacking. We performed BioID in biological triplicates on parental Flp-In T-REx HEK293 cells as well as cell lines stably expressing the BirA*-FLAG tag fused to either the green fluorescent protein (GFP) or a nuclear localization signal (NLS). By immunofluorescence, these controls generated distinct staining patterns of biotinylated proteins (Figure 2A) that we concluded might help us better characterize the effect of the subcellular localization of the BirA* enzyme on the background detected by the approach. We also reasoned that combining controls with different subcellular localization may provide a more complete reference for statistical analysis of the BioID data. Following data acquisition on a TripleTOF 5600 mass spectrometer, the resulting 9 files were processed through the iProphet pipeline; see Materials and Methods for details. In total, 571 proteins were identified with at least two unique peptides in parental Flp-In T-REx HEK293 cells. The relative abundance list (ordered by decreasing spectral counts) was dominated by the metabolic enzymes ACACA, PC, MCCC1, MCCC2 and PCCA that either possess biotin carboxylase activity or are known to be endogenously biotinylated. These proteins were also the most abundant contaminants observed in the BioID purifications from cells expressing either the NLS-BirA*-FLAG or the GFP-BirA*-FLAG constructs. Overall, exogenous expression of a BirA* moiety increased the number of proteins detected following the streptavidin purification to 1448 proteins with the NLS-BirA*-FLAG construct and 1534 proteins with GFP-BirA*-FLAG (Figure 2B). Most proteins identified with high spectral counts were found at similar levels in both GFP- and NLS-BirA* tagged controls, though a number of proteins were identified preferentially with only one type of controls (Figure 2B and C). To ensure that triplicate analysis of our control cell lines was sufficient for proper modeling of the background, we acquired 12 BioID biological replicates for the GFP-BirA*-FLAG control. By plotting the total number of proteins detected with at least 2 unique peptides and an iProphet FDR rate of 5%, we observed that we neared saturation in the number of proteins detected after three biological replicates (86% of proteins detected; supplementary Figure 1). Thus, our initial characterization of BioID background suggests that multiple

types of controls may help to model the abundance of non-bait specific biotinylated proteins more efficiently than using additional replicates of the same type of controls.

Efficient purification of histone interaction partners by BioID

To further explore the BioID approach, we next turned to the analysis of two core histone proteins (H2B and H3) tagged with a C-terminal BirA*-FLAG tag. BioID analysis of both histone cell lines was performed in biological duplicates and showed a reproducibility that was in-line with AP-MS (supplementary Figure 2) [3]. The resulting mass spectrometry results were analyzed using the SAINTexpress algorithm [24]. SAINTexpress enables the “compression” of controls to increase the stringency of data analysis by performing the analysis against the n highest values within a given control dataset (this strategy was recently discussed within the context of the Contaminant Repository for Affinity Purification; [28]). Initially, we used 9 controls (3 empty Flp-In T-REx HEK293, 3 GFP-BirA*-FLAG and 3 NLS-BirA*-FLAG) and compressed them down to 3. We reasoned that this would provide us with a worst case scenario where each type of control would contribute to its fullest. This resulted in the detection of 210 significant interaction partners for H2B and 90 for H3 at a SAINT FDR of 1% (Figure 3A). Omitting the control compression increases the number of significant interaction partners to 309 and 133 for histone H2B and H3, respectively (Figure 3A, supplementary Figure 3). Using 9 GFP controls only for SAINT analysis further relaxed the stringency of the analysis and increased the number of significant interaction partners to 385 (H2B) and 183 (H3). Consequently, we decided to use the most stringent set of controls and SAINT compression parameters for the remainder of our analysis.

Compared to our recent analysis of histone proteins using affinity purification coupled to mass spectrometry (AP-MS) [3], the BioID approach produced a much larger set of significant interaction partners for histones H2B and H3 (Figure 3B). In fact, a surprisingly low number of interaction partners were identified by both BioID and AP-MS for both histones tested (Figure 3B). A larger fraction of interactions were common to both H2B and H3 by BioID than was detected by AP-MS (supplementary Figure 4), consistent with the fact that BioID identifies the environment of a bait in addition to its direct interaction partners. GO term analysis of the significant interaction partners identified by both techniques did further highlight distinct biological processes preferably recovered by BioID or AP-MS (Figure 3C). For instance, we detected numerous interaction partners involved in mitosis associated with both H2B and H3 by BioID but not AP-MS. This may be due to the fact that the incubation with biotin is performed for 24 hours prior to cell harvesting, allowing most cells to progress through a complete cell cycle. Conversely, cells prepared for AP-MS were asynchronous and contained only a small percentage of cells in M phase. This suggests that BioID may be more amenable to the identification of transient or temporally-regulated interactions than AP-MS from asynchronous cells. We note that, while not addressed here, it is possible to synchronize cells prior to AP-MS [29–31] which would allow to enrich interactions present at specific cell cycle stages. For significantly enriched GO terms in common between AP-MS and BioID, such as chromatin organization and response to DNA damage, we observed that different prey proteins contributed to the GO term enrichment (supplementary Figure 5A and B).

To more easily visualize the overlap between the BioID and AP-MS results for histones H2B and H3, we generated a dotplot displaying the absolute spectral counts of a given prey concurrently with its relative abundance across all samples (supplementary Fig 6). When we focused on histone chaperones and other known chromatin-associated proteins, we observed some clear differences in the purification efficiency of different preys by BioID or AP-MS (Fig 4A). For example, the CAF1 protein complex, composed of CHAF1A and CHAF1B, which participates in the replication-coupled deposition of histone H3 and H4 [32], was identified as a significant interaction partner by both BioID and AP-MS but with much higher spectral counts in BioID (Figure 4A). Looking at other histone chaperone complexes detected, we found that some were favored by AP-MS (NAP1L1/NAP1L4 association with H2B), BioID (CHAF1A/CHAF1B association with H3) or both (ASF1A/ASF1B found in similar amounts with H3). No obvious purification bias for histone chaperones toward BioID or AP-MS was observed. By using the PaxDB protein database [33] as a proxy for protein abundance, we did observe that prey proteins detected by BioID have statistically significant lower abundance than those observed with AP-MS (Figure 4B). While this may be a consequence of the novelty of BioID (i.e. BioID MS data has not contributed significantly to the PaxDB database yet while standard AP-MS has), it also suggests that BioID facilitates the isolation of proteins difficult to purify by established biochemical approaches. Taken together, this dataset suggests that while both purification approaches succeed in generating interactome maps for histones H2B and H3, performing both methods allowed us to access an expanded protein complex space.

Characterization of the mediator complex by AP-MS and BioID

We next addressed whether BioID was amenable to the study of large and well-structured chromatin-associated protein complexes. We performed analysis of the 1.2 MDa mediator complex, which contains over 30 subunits [34]. The mediator complex is critical in the regulation of transcription by RNA polymerase II (RNAPII) and has been implicated in a wide array of functions such as mRNA splicing and non-coding RNA activation [35, 36]. We selected three subunits of mediator located in the middle (MED4), head (MED20) and tail (MED23) modules of this complex (Figure 5A) for BioID and AP-MS. Stable cell lines expressing either the 3XFLAG or BirA*-FLAG tag were generated in Flp-In T-REx HEK293 cells and shown to express the recombinant proteins at similar levels to the GFP-BirA*-FLAG and NLS-BirA*-FLAG tagged controls, with the notable exception of the MED20-BirA*-FLAG cell line which showed no detectable expression (Figure 5B). Furthermore, the cell lines containing a BirA* tag displayed extensive protein biotinylation while the ones with 3XFLAG tags did not (Figure 5B). Biological replicates of each cell line were harvested and subjected to BioID and AP-MS and the resulting data analyzed using SAINTexpress. In total, 212 and 236 proteins were detected with at least one mediator subunit by BioID and AP-MS, respectively (supplementary table 4). Both BioID and AP-MS were successful at purifying the majority of the known mediator complex components, though differences were observed (see below). The spectral counts for the bait in comparison to other *bona fide* interactors (here mediator complex components) were more similar in the BioID approach, while the bait counts tend to be much higher than the preys in FLAG AP-MS (supplementary Table 5), perhaps because of dissociation of the complex through purification.

Looking globally at the mediator complex interaction partners identified with individual subunits by BioID and AP-MS, we found distinct prey profiles for each sample analyzed (Figure 6A). Our capacity to detect most subunits of mediator was dependent on both the subunits and the purification method used. For example, BioID of BirA*-FLAG-MED23 successfully identified 25 mediator subunits while AP-MS of 3XFLAG-MED23 only detected 7 significant interaction partners (Figure 6B). On the other hand, purifications of MED4 resulted in similar results for BioID and AP-MS with 20 and 28 mediator subunits identified, respectively (Figure 6B). When mediator subunits were excluded, little overlap was observed between the BioID purifications (Figure 6C). While AP-MS showed a larger overlap between subunits purifications, less than 50% of any given interaction partners were shared with the other two subunits tested here.

Localization of MED4 mediator subunits to the centrosome

One of the dominating features of the BioID data obtained for the three mediator subunits studied here is the abundance of centrosomal proteins identified with MED4 (Figure 7A). The centrosome serves as the microtubule-organizing centre of the cell, enabling key cellular functions, such as locomotion and division, to proceed properly [37]. Centrosomes are large protein structures, with over 100 proteins reported to possess a centrosomal localization [38]. Surprisingly, numerous structural constituents of the centrosome such as PCM1, AZI1, OFD1 and PCNT were detected at high levels with MED4 by BioID and AP-MS but not with MED20 and MED23 (Figure 7A). As the mediator complex has not been previously linked to the centrosome, we sought to validate this finding. To do so, Flp-In T-REx HEK293 cells stably expressing BirA*-FLAG tagged MED4 or MED23 were used for immunofluorescence. In interphase cells, we observed BirA*-FLAG-MED4 to localize to a single dense perinuclear protein foci consistent with the centrosome (Figure 7B). BirA*-FLAG-MED23, on the other hand, displayed a nuclear localization (Figure 7B). Co-localization experiments using the well-established centrosome markers pericentrin and centrin further confirmed this finding (Figure 7C). In addition, cells undergoing division showed two 3XFLAG-MED4 foci on either side of the aligned chromosomes, co-localizing with centrin or pericentrin, also consistent with centrosomal localization throughout the cell cycle (Figure 7C). Together, this supports a novel role for MED4 at centrosomes, though more thorough studies will need to be performed to define its role at this location.

Conclusions

The choice of an approach for mapping protein-protein interactions is in some way akin to a Faustian bargain. While protein-protein interaction mapping, independently of the approach used, does provide researchers with key biological insights, it also biases their view of a protein interactome toward a particular subset of interaction partners. This has been observed countless times, manifesting itself in incomplete overlaps between the interactions captured by different approaches [39, 40], even after correcting for technical issues and controlling false positive detection rates [41]. In this context, and in light of the previous BioID studies [11, 13–17], the partial overlap detected is therefore not completely surprising. Here, we found that BioID generally produced larger interactomes than AP-MS, despite using the same bait sequence, cell type, mass spectrometer detector and statistical

analysis. In addition, we observed that BioID identified prey proteins of significantly lower abundance than those observed with AP-MS. This being said, both BioID and FLAG AP-MS were able to recapitulate a fraction of protein-protein interactions previously identified using other biochemical approaches, and provided complementary information for a given protein bait. BioID is therefore an interesting addition to current assays mapping protein-protein interactions for nuclear and chromatin-associated proteins. As a note of caution, however, we also note the strong impact of the selection of the type of controls and scoring mechanism for the BioID data: this emphasizes the need for strong experimental design when performing this approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank Nicole St-Denis, Amber Couzens and other members of the Gingras lab for helpful discussions and help with editing. The NLS-BirA*-FLAG construct was a generous gift of Catherine Brun and Daniel Durocher (Lunenfeld-Tanenbaum Research Institute). This work was supported by funding from the Canadian Institutes of Health Research (CIHR) to A.-C.G. (MOP 123322) and from the National Institutes of Health to A.-C.G. (5R01GM94231). A.-C.G. holds the Canada Research Chair in Functional Proteomics and the Lea Reichmann Chair in Cancer Proteomics. J.-P.L. was supported by a post-doctoral fellowship from the CIHR and by a TD Bank Health Research Fellowship at the Lunenfeld-Tanenbaum Research Institute. J.D.R.K. was supported by a postdoctoral fellowship from the Canadian Heart and Stroke Foundation. C.G. was supported by a master award Frederick Banting and Charles Best Canada Graduate Scholarships from CIHR.

References

1. Lambert JP, Pawson T, Gingras AC. Mapping physical interactions within chromatin by proteomic approaches. *Proteomics*. 2012; 12:1609–22. [PubMed: 22611019]
2. Kean MJ, Couzens AL, Gingras AC. Mass spectrometry approaches to study mammalian kinase and phosphatase associated proteins. *Methods*. 2012; 57:400–8. [PubMed: 22710030]
3. Lambert JP, Tucholska M, Pawson T, Gingras AC. Incorporating DNA shearing in standard affinity purification allows simultaneous identification of both soluble and chromatin-bound interaction partners. *Journal of proteomics*. 2014; 100:55–9. [PubMed: 24412199]
4. Mak AB, Ni Z, Hewel JA, Chen GI, Zhong G, Karamboulas K, et al. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol Cell Proteomics*. 2010; 9:811–23. [PubMed: 20305087]
5. Marcon E, Ni Z, Pu S, Turinsky AL, Trimble SS, Olsen JB, et al. Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. *Cell reports*. 2014; 8:297–310. [PubMed: 24981860]
6. Lambert JP, Mitchell L, Rudner A, Baetz K, Figeys D. A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol Cell Proteomics*. 2009; 8:870–82. [PubMed: 19106085]
7. Lambert JP, Fillingham J, Siahbazi M, Greenblatt J, Baetz K, Figeys D. Defining the budding yeast chromatin-associated interactome. *Molecular systems biology*. 2010; 6:448. [PubMed: 21179020]
8. Gingras AC, Raught B. Beyond hairballs: The use of quantitative mass spectrometry data to understand protein-protein interactions. *FEBS letters*. 2012; 586:2723–31. [PubMed: 22710165]
9. Stengel F, Aebersold R, Robinson CV. Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Mol Cell Proteomics*. 2012; 11:R111 014027. [PubMed: 22180098]

10. Glatter T, Schittenhelm RB, Rinner O, Roguska K, Wepf A, Junger MA, et al. Modularity and hormone sensitivity of the *Drosophila melanogaster* insulin receptor/target of rapamycin interaction proteome. *Molecular systems biology*. 2011; 7:547. [PubMed: 22068330]
11. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology*. 2012; 196:801–10. [PubMed: 22412018]
12. Kwon K, Beckett D. Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein science: a publication of the Protein Society*. 2000; 9:1530–9. [PubMed: 10975574]
13. Firat-Karalar EN, Rauniyar N, Yates JR 3rd, Stearns T. Proximity interactions among centrosome components identify regulators of centriole duplication. *Curr Biol*. 2014; 24:664–70. [PubMed: 24613305]
14. Comartin D, Gupta GD, Fussner E, Coyaud E, Hasegan M, Archinti M, et al. CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Curr Biol*. 2013; 23:1360–6. [PubMed: 23810536]
15. Morriswood B, Havlicek K, Demmel L, Yavuz S, Sealey-Cardona M, Vidilaseris K, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryotic cell*. 2013; 12:356–67. [PubMed: 23264645]
16. Kim DI, Kc B, Zhu W, Motamedchaboki K, Doye V, Roux KJ. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E2453–61. [PubMed: 24927568]
17. Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, et al. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Science signaling*. 2013; 6:rs15. [PubMed: 24255178]
18. Lamesch P, Li N, Milstein S, Fan C, Hao T, Szabo G, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics*. 2007; 89:307–15. [PubMed: 17207965]
19. Yang X, Boehm JS, Salehi-Ashtiani K, Hao T, Shen Y, Lubonja R, et al. A public genome-scale lentiviral expression library of human ORFs. *Nat Methods*. 2011; 8:659–61. [PubMed: 21706014]
20. Olhovskiy M, Williton K, Dai AY, Pasculescu A, Lee JP, Goudreaux M, et al. OpenFreezer: a reagent information management software system. *Nat Methods*. 2011; 8:612–3. [PubMed: 21799493]
21. Liu G, Zhang J, Larsen B, Stark C, Breitkreutz A, Lin ZY, et al. ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nature biotechnology*. 2010; 28:1015–7.
22. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, et al. A guided tour of the Trans-Proteomic Pipeline. *Proteomics*. 2010; 10:1150–9. [PubMed: 20101611]
23. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics*. 2011; 10:M111 007690. [PubMed: 21876204]
24. Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras AC, Choi H. SAINTexpress: improvements and additional features in Significance Analysis of INteractome software. *Journal of proteomics*. 2014; 100:37–43. [PubMed: 24513533]
25. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009; 4:44–57.
26. Knight JDR, Liu G, Zhang JP, Pasculescu A, Choi H, Gingras AC. A web-tool for visualizing quantitative protein-protein interaction data. *Proteomics*. 2014 submitted.
27. Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: a web tool for generation of box plots. *Nat Methods*. 2014; 11:121–2. [PubMed: 24481215]
28. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*. 2013; 10:730–6. [PubMed: 23921808]
29. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Heriche JK, Sykora MM, et al. Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science*. 2010; 328:593–9. [PubMed: 20360068]

30. Pijnappel WP, Kolkman A, Baltissen MP, Heck A Jr, Timmers HM. Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. *Proteome Sci.* 2009; 7:46. [PubMed: 20034391]
31. Ma H, McLean JR, Chao LF, Mana-Capelli S, Paramasivam M, Hagstrom KA, et al. A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms. *Mol Cell Proteomics.* 2012; 11:501–11. [PubMed: 22474084]
32. Loyola A, Almouzni G. Histone chaperones, a supporting role in the limelight. *Biochimica et biophysica acta.* 2004; 1677:3–11. [PubMed: 15020040]
33. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics.* 2012; 11:492–500. [PubMed: 22535208]
34. Tsai KL, Sato S, Tomomori-Sato C, Conaway RC, Conaway JW, Asturias FJ. A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature structural & molecular biology.* 2013; 20:611–9.
35. Conaway RC, Conaway JW. The Mediator complex and transcription elongation. *Biochimica et biophysica acta.* 2013; 1829:69–75. [PubMed: 22983086]
36. Schiano C, Casamassimi A, Rienzo M, de Nigris F, Sommese L, Napoli C. Involvement of Mediator complex in malignancy. *Biochimica et biophysica acta.* 2014; 1845:66–83. [PubMed: 24342527]
37. Bornens M. The centrosome in cells and organisms. *Science.* 2012; 335:422–6. [PubMed: 22282802]
38. Nigg EA, Raff JW. Centrioles, centrosomes, and cilia in health and disease. *Cell.* 2009; 139:663–78. [PubMed: 19914163]
39. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature.* 2002; 417:399–403. [PubMed: 12000970]
40. Bader JS, Chaudhuri A, Rothberg JM, Chant J. Gaining confidence in high-throughput protein interaction networks. *Nature biotechnology.* 2004; 22:78–85.
41. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, et al. An empirical framework for binary interactome mapping. *Nat Methods.* 2009; 6:83–90. [PubMed: 19060904]

Significance

This manuscript describes the application of BioID, a proximity biotinylation approach, to chromatin-associated proteins, namely core histones and members of the mediator complex. We observed that BioID was successful at identifying known interaction partners for the baits tested, but also allowed novel putative interaction partners to be identified. By performing a detailed comparison of BioID versus a standard method for interactome mapping (affinity purification coupled to mass spectrometry, AP-MS), we show that the approaches were complementary, allowing for purification of different interaction partners. These interaction partners were different in the biological processes they are associated with, but also in their abundance. BioID represents a significant technical development in the field of chromatin research by expanding the search space for interactome mapping beyond what is possible with AP-MS.

Highlights

- BioID is effective for the interactome mapping of chromatin associated proteins.
- BioID and AP-MS produce largely distinct interactomes.
- The selection of the subunit to be tagged affects interactome mapping results.

Figure 1. Overview of BioID and AP-MS approaches for interactome mapping. See text and methods for details.

Figure 2.

The BioID background is affected by exogenous BirA* expression and localization. (A) T-REx Flp-In HEK293 cell lines stably expressing GFP-BirA*-FLAG, NLS-BirA*-FLAG or no constructs (empty) grown on poly-L-lysine coated cover slips. Cells were fixed with formaldehyde and stained for biotinylated proteins, FLAG-tagged proteins (baits) and DAPI (nucleus). (B) Venn diagram showing the overlap of proteins identified by BioID from T-REx Flp-In HEK293 cells stably expressing GFP-BirA*-FLAG, NLS-BirA*-FLAG or no constructs (empty). (C) Sum of spectral counts across three replicates for proteins identified in BioID analysis from T-REx Flp-In HEK293 cells stably expressing GFP-BirA*-FLAG or NLS-BirA*-FLAG. The diagonal line maps to equal spectral counts across the GFP-BirA*-FLAG and NLS-BirA*-FLAG samples.

Figure 3.

BioID is successful at mapping interactomes of histones H2B and H3. (A) Venn diagram of significant interaction partners detected by BioID analysis of histone H2B and H3 after running SAINTexpress using different types of background controls and compression parameters. (B) Venn diagram of significant interaction partners detected by BioID and AP-MS (data generated in [3]) analyses of histones H2B and H3. (C) Functional annotation of the significant interaction partners for H2B and H3; the protein count for biological processes (GO FAT) with Benjamini-Hochberg FDR correction of 0.01 or smaller is mapped. Proteins contributing to GO term GO:0051276 (*, chromatin organization) and GO:0006974 (**, response to DNA damage stimulus) are shown in supplementary Figure 2.

Figure 4.

BioID and AP-MS show distinct profiles of interaction partners for histones H2B and H3. (A) Dotplot of selected interaction partners identified by BioID and AP-MS. Node color represents the absolute spectral count sum (capped at 50 here for display purposes), the node edge color corresponds to the SAINTexpress FDR value and the node size displays the relative abundance of a given prey across the four samples compared. Histone chaperones are shown in bold italics. Complete dotplot representation can be seen in supplementary Figure 6. (B) Box plot representation of PaxDB abundance for H2B and H3 interaction partners detected by BioID and AP-MS. Interaction partners of histones H2B and H3 detected by BioID are of significantly (paired 2 t-tailed t-test < 0.05) lower abundance than those detected by AP-MS.

Figure 5.

Expression levels of 3XFLAG and BirA*-FLAG tagged mediator subunits in T-REx Flp-In HEK293 stable cell lines. (A) Depiction of the four modules of the mediator complex and of their protein constituents. (B) Expression level of controls and tagged mediator subunits as assessed by Western blot. Lysates from the indicated cell lines were resolved in 4–15% gradient gels and probed with anti-FLAG M2 or anti-tubulin antibodies. Subsequently, blots were reprobed with streptavidin conjugated to HRP to detect the extent of protein biotinylation. Please note that the T-REx Flp-In HEK293 stable cell line expressing MED20-BirA*-FLAG shows no visible expression or additional protein biotinylation, despite being detected (albeit at low levels) by MS.

Figure 6. Analysis of mediator by BioID and AP-MS reveals distinct interactome profiles for MED4, MED20 and MED23. (A) Hierarchical clustering (Pearson correlation, complete linkage clustering) of SAINTexpress output from AP-MS and BioID data obtained for MED4, MED20 and MED23. (B) Highlight of the mediator complex (box #1) observed following SAINTexpress analysis. Mediator complex subunits were colored as per legend. Highlighted box #2 is associated with Figure 7A. Please note that bait proteins are not displayed in the heat map presented in panel A and B. (C) Summary of the overlap between the interaction partners of the three mediator subunits studied by BioID and AP-MS, excluding mediator subunits themselves.

Figure 7.

MED4 localizes to the centrosome throughout the cell cycle. (A) BioID and AP-MS analyses of MED4 reveal significant association with centrosomal proteins. Highlighted box #2 from the heat map presented in Figure 6 is shown. (B) BirA*-FLAG tagged MED4 shows strong immunofluorescence staining consistent with centrosomal localization in interphase cells (top panels). BirA*-FLAG tagged MED23 is only observed in the nucleus but not at centrosomes (bottom panels). 3XFLAG-MED4 co-localizes with two centrosomal markers, pericentrin (C) and centrin (D), by immunofluorescence in interphase cells (top panels) and in cells undergoing division (bottom panels).