Novel methods for studying multiprotein complexes in vivo

Virja Mehta and Laura Trinkle-Mulcahy*

Address: Department of Cellular and Molecular Medicine, Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, ON K1H 8M5, Canada

* Corresponding author: Laura Trinkle-Mulcahy (ltrinkle@uottawa.ca)

F1000Prime Reports 2013, 5:30 (doi:10.12703/P5-30)

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License (http://creativecommons.org/licenses/by-nc/3.0/legalcode), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. You may not use this work for commercial purposes.

The electronic version of this article is the complete one and can be found at: http://f1000.com/prime/reports/m/5/30

Abstract

The current consensus is that the majority of proteins act in concert in the cell, as homo- and heteromeric complexes of two or more proteins that carry out discrete biological functions. A wide range of genomic, proteomic, biochemical, structural and biophotonic techniques have been employed over the years to study the protein-protein interactions that define complexes, with the end goal of producing a spatiotemporal map of these modular functional units throughout the cell. Recent advances in the analysis of in vivo complexes have greatly improved structural, functional and temporal resolution, and this review highlights novel approaches ranging from proximity-dependent labeling and cross-linking/mass spectrometry through pulse-chase epitope labeling and targeted protein degradation.

Structural analyses of multiprotein complexes

“No man is an island entire of itself” (John Donne).

With the assembly of proteins into functional complexes thought to underlie most, if not all, biological processes, characterization of these structures is a key goal in cell biology. An initial step is identification of complex members, both stably- and transiently-associated, and their intra-complex interactions. The current method of choice for interactome analyses is affinity purification followed by mass spectrometry (AP-MS; for review see [1-3]), and the development of quantitative methods has enabled resolution of the components of large multiprotein complexes and provided information about their stoichiometry [4-7]. What this type of approach on its own does not provide, however, is information about the topological structure of the complex and the functional significance of each member. Furthermore, it traditionally involves breaking open cells to extract complexes for analysis, a process that can be disruptive to the underlying protein-protein associations.

BioID is a recently developed technique that complements traditional AP/MS-based interactome mapping by highlighting intracellular protein “neighbours” in vivo, using a proximity labelling/affinity purification strategy [8]. It was inspired by the DamID method utilized to detect DNA-protein interactions via methylation of DNA sequences proximal to a DNA binding protein fused to Dam methylase [9]. In the BioID approach, a promiscuous prokaryotic biotin protein ligase (BirA*) is fused to the protein of interest. When expressed in cells, the fusion protein will biotinylate proteins with which it comes into close proximity, such as direct binding partners and neighbouring proteins in multiprotein complexes (Fig. 1A). Importantly, it has the ability to capture both stable interactions and transient or weak interactions. Biotinylated proteins can be isolated by affinity purification, using a streptavidin agarose matrix, and identified by MS. Caveats include the inability to distinguish direct vs. indirect interactors (similar to AP/MS) and the unknown activity radius of BirA*, which would define the resolution of this technique. As an initial non-biased screen for in vivo associations, however, a major strength is the accessibility of this method to a wide range of researchers, in that it requires only standard molecular and cell biology techniques and access to proteomics services.
A. In the BioID approach, fusion of a promiscuous *E. coli* biotin protein ligase (BirA*) to the protein of interest promotes biotinylation of near-neighbour proteins *in vivo*. These biotinylated proteins can then be captured by affinity purification and identified by mass spectrometry. B. The combination of cross-linking with mass spectrometry can provide information about protein-protein interactions and multiprotein complex architecture. Complexes (either *in vivo* or affinity purified) are treated with a bi-functional cross-linking reagent that creates a covalent link between adjacent regions of polypeptide chains. These links can be intra-chain (within the same protein; green) or inter-chain (within neighbouring proteins; red). Proteolytic digests are then analyzed by LC-MS/MS to identify cross-linked peptides, which in turn provide structural information about the protein complexes. C. A non-radioactive translation-controlled pulse-chase system that enables spatiotemporal monitoring of the biogenesis of multiprotein complexes. Left panel: Cells transformed with a plasmid encoding the gene of interest (with a C-terminal affinity tag) downstream of an HA tag and amber stop codon (UAG) only synthesize the HA peptide due to translational termination at this premature stop codon. Middle panel: Co-expression of an engineered orthogonal pair of amber suppressor tRNA_{Ome-Tyr} and tRNA-synthetase allows incorporation of the unnatural amino acid O-methyl tyrosine (Ome-Tyr) and suppression of the UAG stop codon, leading to expression of the full HA-protein-affinity tag construct. Addition of Ome-Tyr to the media thus induces a translational pulse of tagged protein expression. Right panel: Due to a tetracycline-regulatable riboswitch engineered into the 5' UTR of the HA-UAG-gene-affinity tag plasmid, synthesis can then be inhibited at any time by addition of tetracycline (translational chase). Affinity purification of the tagged protein and interactome mapping at different time points following the pulse can be used to probe changes in complex composition.
Higher resolution probing of the topology of multi-protein complexes, both in vitro and in vivo, has been enabled by the coupling of classic chemical cross-linking techniques with recent advances in mass spectrometer instrumentation, proteomic methodologies and bioinformatics (for review see [10-13]). Cross-linking provides proximity information, revealing not only which proteins are getting cross-linked but also at which sites the cross-linking takes place (Fig. 1B). The combination of cross-linking with mass spectrometry has been utilized, both on its own and in combination with other structural analysis methods to probe the architecture of complexes such as ribosomes [14], proteasomes [15], RNA polymerase II–TFIIF [16], and the protein phosphatase 2A (PP2A) phosphatase network [17]. Although challenges for cross-linking/MS include the low abundance of cross-linked peptides and the complexity of the fragmentation spectra, these are being addressed by the development of more efficient, affinity tag-linked cross-linking reagents and more powerful data analysis and quantitation software [12,18]. Incorporating quantitative measurements further extends the ability of this approach to analyze the dynamic assembly/disassembly and functional composition of multiprotein complexes [19].

The biological importance of the assembly order of multiprotein complexes was recently demonstrated by the analysis of evolutionary gene fusion events in a large number of species, which revealed that protein complexes are under evolutionary selection to assemble via ordered pathways [20]. Mapping a dynamic assembly process using proteomic approaches requires sufficient temporal resolution, which can be provided by a recently developed protein translation-controlled pulse-chase system (Fig. 1C; [21]). In this approach, a pulse of de novo synthesis of a tagged protein is followed by a time course of affinity purification and interactome mapping to reveal dynamic changes in the composition of complexes to which it is targeted. High time resolution is achieved by controlling protein expression at the level of translation. Cells are transformed both with a plasmid encoding the gene of interest (flanked by N- and C-terminal affinity tags) with an in-frame amber stop codon (UAG) inserted just after the N-terminal tag, and with a plasmid encoding an engineered orthogonal pair of amber suppressor tRNA\(^{\text{Ome-Tyr}}\) and tRNA synthetase. In the absence of Ome-Tyr, only the N-terminal tag is synthesized because translation halts at the UAG (Fig. 1C). Addition of Ome-Tyr to the media allows cells to incorporate this unnatural amino acid at the UAG, leading to translation of the full-length fusion protein. A tc-apta riboswitch engineered into the 5’ end of the transcript allows the “pulse” of de novo protein expression to be rapidly shut down upon addition of tetracycline, which binds the riboswitch and interferes with translation initiation. Although this pulse-chase method was developed in yeast, incorporation of unnatural amino acids via orthogonal amber suppressor tRNA/tRNA synthetase pairs (for review see [22]) extends its use to other biological systems, including Drosophila [23] and mammalian cells [24], highlighting the potentially broad applicability of this technique for the spatiotemporal analysis of multiprotein complex assembly.

**Functional analyses of multiprotein complexes**

Valuable clues to the physiological function of a protein can be obtained by observing the downstream phenotypic effects of removing it from cells or organisms (Fig. 2A). At the DNA level, powerful knockout and mutagenesis approaches based on homologous recombination and zinc finger nucleases have enabled targeted deletion/mutation of specific genes in model systems, including yeast, mammalian cells and mice (for review see [25-27]). At the RNA level, post-transcriptional gene silencing techniques based on RNA interference (RNAi; for review see [28,29]) permit the knockdown of specific proteins via targeted degradation of their mRNA. A caveat to these approaches is that both lie upstream of the protein itself, incorporated within its functional multiprotein complex (es), and thus necessitate a time delay until protein levels are reduced (which will vary based on protein turnover rates) or limit studies into the effects of permanent removal of the protein.

In contrast, targeted destruction offers the ability to study the acute effects of the immediate removal of a protein. One in vitro approach is the spatially and temporally defined photo-ablation method Chromophore-Assisted Light Inactivation (CALI; for review see [30]), in which the target protein is fused to a fluorophore, such as KillerRed, that produces substantial amounts of reactive oxygen species (ROS) upon absorption of light at a particular wavelength. The end result is destruction of the target protein in the region of interest, although caveats include inadvertent protein cross-linking and inactivation of proteins beyond the immediate target by diffusion of the ROS. Alternative in vivo methods that take advantage of the endogenous ubiquitin-mediated protein degradation pathway [31] have been developed, and two recent protein knockout techniques, based on the SKP1/CUL1/F-box (SCF) protein ubiquitin ligase complex, extend both the in vivo applicability and time resolution of this type of approach.

By fusing an F-box protein derived from Drosophila (Slmb) with a single-domain antibody fragment (vhhGFP4) that recognizes green fluorescent protein
(GFP) and related derivatives, Caussinus and colleagues created a genetically encoded method, which they call deGradFP, that enables the direct depletion of target proteins fused to these fluorophores [32]. Recruitment of endogenous SCF complex members by Slmb leads to ubiquitination of the fusion protein and its targeted degradation by the proteasome (Fig. 2B), which can be monitored via fluorescence imaging. Although a major strength of this approach is its potential applicability to any GFP-tagged construct, the authors noted the failure of deGradFP to induce degradation of both free GFP and a particular protein incorporated into adherens junctions. While this suggests possible structural and accessibility limitations, it will be easier to judge the extent and implications of this as the technique is applied to a wider range of substrates.
Auxin-induced degradation, based on a unique ligand-induced degradation pathway in plants, is a related approach that can further increase the time resolution of protein depletion and is fully reversible [33,34]. In this system, a target protein is expressed as a fusion with an auxin-inducible degron (AID) in cells that also exogenously express the plant F-box protein Transport Inhibitor Response 1 (TIR1). Auxin hormones such as indole-3-acetic acid (IAA) promote the interaction between AID-containing proteins and TIR1, which can in turn recruit endogenous SCF proteins and promote ubiquitination and proteasome-mediated degradation of the target protein (Fig. 2C). Depletion is thus inducible, rapid and complete, and fully reversible upon removal of IAA. In this study, the authors demonstrate targeted depletion of a wide range of substrates localized to different regions of mammalian cells, including nucleosomal histones, centromere-, centromere- and telomere-associated proteins and cytoplasmic cyclin B1. Degradation could be readily monitored by fluorescence imaging, due to an additional GFP (or related derivative) tag added to the fusion proteins. As with any fusion protein, the addition of tags (in this case 25 kDa AID plus 27 kDa GFP) has the potential to affect localization and function, and thus proper in vivo behavior must first be demonstrated. Furthermore, a limitation of both deGradFP and AID-induced degradation is that they cannot control the stability of endogenous untagged proteins. The continued refinement of homologous recombination strategies in various model systems does, however, offer the potential to extend these technologies by genetically encoding the tags in frame with endogenous genes.

Conclusion

Taken together, these novel methods for dissecting multiprotein complexes in vivo offer unprecedented sensitivity and spatiotemporal resolution, which is a large step toward the ultimate goal of mapping functional multiprotein complexes throughout the cell under a variety of conditions. Importantly, most do not require specialized knowledge or equipment and are therefore accessible to a wide range of researchers. They can also be adapted to different model systems and further modified to increase their resolution and applicability.

Abbreviations

AID, auxin-inducible degron; AP-MS, affinity purification-mass spectrometry; CALI, chromophore-assisted light inactivation; GFP, green fluorescent protein; IAA, indole-3-acetic acid; LC-MS/MS, Liquid chromatography-tandem mass spectrometry; Ome-Tyr, O-methyl tyrosine; PP2A, protein phosphatase 2A; RNAi, RNA interference; SCF, SKP1/CUL1/F-box protein ubiquitin ligase; TIR1, Transport Inhibitor Response 1.

Disclosure

The authors declare that they have no disclosures.

Acknowledgments

Virja Mehta is supported by a Frederick Banting and Charles Best Canada Graduate Scholarship from the Canadian Institutes of Health Research. Laura Trinkle-Mulcahy’s laboratory is supported by the Canadian Cancer Society Research Institute and Terry Fox Foundation and by the Natural Sciences and Engineering Research Council of Canada. Laura Trinkle-Mulcahy holds a Canadian Institutes of Health Research New Investigator Salary Award.

References

1. Trinkle-Mulcahy L: Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry. Proteomics 2012, 12:e1623-38.
2. Gingras A-C, Gstaiger M, Raught B, Aebersold R: Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 2007, 8:645-54.
3. Vermeulen M, Hubner NC, Mann M: High confidence determination of specific protein–protein interactions using quantitative mass spectrometry. Current Opinion in Biotechnology 2008, 19:331-7.
4. Kaake R, Wang X, Huang L: Profiling of Protein Interaction Networks of Protein Complexes Using Affinity Purification and Quantitative Mass Spectrometry. Molecular and Cellular Proteomics 2010, 9:1650-65.
5. Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert F-M, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H, Lamond A: Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol 2008, 183:223-39.
6. Schmidt C, Lenz C, Grote M, Luhrmann R, Urlaub H: Determination of Protein Stoichiometry within Protein Complexes Using Absolute Quantification and Multiple Reaction Monitoring. Anal Chem 2010, 82:2784-96.
7. Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M: Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Research 2012, 41:e28-e28.
8. Roux Kj, Kim DI, Raida M, Burke B: A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol 2012, 196:801-10.
9. Steensel BV, Henikoff S: Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. Nat Biotechnol 2000, 18:424-8.
10. Chavez JD, Weisbrod CR, Zheng C, Eng JK, Bruce JE: In VivoProtein Interaction Network Identified with a Novel
Real-Time Cross-Linked Peptide Identification Strategy. J Proteome Res 2013, 12:1569-79.

11. Stengel F, Aebersold R, Robinson CV: Joining Forces: Integrating Proteomics and Cross-linking with the Mass Spectrometry of Intact Complexes. Mol Cell Proteomics 2012, 11:1-13.

12. Petrotenko EV, Borchers CH: Cross-linking combined with mass spectrometry for structural proteomics. Mass Spectrom Rev 2010, 29:862-76.

13. Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R: Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. Mol Cell Proteomics 2010, 9:1634-49.

14. Lauber MA, Reilly JP: Structural Analysis of a Prokaryotic Ribosome Using a Novel Amidinating Cross-Linker and Mass Spectrometry. J Proteome Res 2011, 10:3604-16.

15. Bohn S, Beck F, Sakata E, Walzthoeni T, Beck M, Aebersold R, Förster F, Baumeister W: Structure of the 26S proteasome from Schizosaccharomyces pombe at subnanometer resolution. Proc Natl Acad Sci USA 2010, 107:20992-7.

16. Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Lariviere L, Bukowski-Wills J-C, Nilges M, Cramer P, Aebersold R: Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J 2010, 29:717-26.

17. Herzog F, Kahraman A, Boebringer D, Mak R, Bracher A, Walzthoeni T, Leitner A, Beck M, Hartl FU, Ban N, Malmstrom L, Aebersold R: Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. Science 2012, 337:1348-52.

18. Luo J, Fishburn J, Hahn S, Ranish J: An Integrated Chemical Cross-linking and Mass Spectrometry Approach to Study Protein Complex Architecture and Function. Mol Cell Proteomics 2012, 11: M111.008318.

19. Fischer L, Chen ZA, Rappsilber J: Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. J Proteomics 2013, 2:120-8.

20. Marsh JA, Hernández H, Hall Z, Ahnert SE, Perica T, Robinson CV, Teichmann SA: Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. Cell 2013, 153:461-70.

21. Stelter P, Kunze R, Radwan M, Thomson E, Thierbach K, Thoms M, Hurt E: Monitoring Spatiotemporal Biogenesis of Macromolecular Assemblies by Pulse-Chase Epitope Labeling. Molecular Cell 2012, 47:788-96.

22. Young TS, Schultz PG: Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. Journal of Biological Chemistry 2010, 285:11039-44.

23. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW: Expanding the genetic code of Drosophila melanogaster. Nat Chem Biol 2012, 8:748-50.

24. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, Lee K-F, Slesinger PA, Wang L: Genetically encoding unnatural amino acids for cellular and neuronal studies. Nat Neurosci 2007, 10:1063-72.

25. Forburg SL: The art and design of genetic screens: yeast. Nat Rev Genet 2001, 2:659-68.

26. Carroll D: Genome Engineering With Zinc-Finger Nucleases. Genetics 2011, 188:773-82.

27. Mallo M: Controlled gene activation and inactivation in the mouse. Frontiers in Bioscience 2006, 300-12.

28. Castel SE, Martienssen RA: RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 2013, 14:100-12.

29. Rao DD, Vorhies JS, Senzer N, Nemunaitis J: siRNA vs. shRNA: Similarities and differences. Advanced Drug Delivery Reviews 2009, 61:746-59.

30. Jacobson K, Rajfur Z, Vitriol E, Hahn K: Chromophore-assisted laser inactivation in cell biology. Trends in Cell Biology 2008, 18:443-50.

31. Zhou P, Bogacki R, McReynolds L, Howley PM: Harnessing the Ubiquitination Machinery to Target the Degradation of Specific Cellular Proteins. Molecular Cell 2000, 6:751-6.

32. Caussinus E, Kanca O, Affolter M: Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 2011, 19:117-21.

33. Holland AJ, Fachinetti D, Han JS, Cleveland DW: Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci USA 2012, 109: E3350-E3357.

34. Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M: An auxin-based degron system for the rapid depletion of proteins in non plant cells. Nat Methods 2009, 6:917-22.