Approaches for Assessing and Discovering Protein Interactions in Cancer

Hisham Mohammed and Jason S. Carroll

Abstract

Significant insight into the function of proteins can be delineated by discovering and characterizing interacting proteins. There are numerous methods for the discovery of unknown associated protein networks, with purification of the bait (the protein of interest) followed by mass spectrometry as a common theme. In recent years, advances have permitted the purification of endogenous proteins and methods for scaling down starting material. As such, approaches for rapid, unbiased identification of protein interactomes are becoming a standard tool in the researchers toolbox, rather than a technique that is only available to specialists. This review will highlight some of the recent technical advances in proteomic-based discovery approaches, the pros and cons of various methods and some of the key findings in cancer-related systems. Mol Cancer Res; 11(11); 1295–302. ©2013 AACR

Introduction

The number of studies that have used protein interaction discovery approaches for identification of associated factors is overwhelming. These tend to fall roughly into two categories: validation of putative or suspected interactions and identification of unknown associated proteins. This review will cover both, but with an emphasis on novel methods for global identification of unknown interactors. We do not aim to provide an exhaustive analysis of all available techniques, but instead will highlight some of the most widely used methods for protein–protein assessment and will describe some recent advances in global protein–protein analyses that are enabling non-experts to identify interactomes of their favorite factor. Where specific biologic examples are used, these only include cancer-related studies.

Direct Assessment of Known Protein–Protein Interactions

The classic method for analysis of protein–protein interactions is the co-immunoprecipitation (co-IP), also known as a "pull down" (1, 2), which constitutes the gold standard in the field. This involves purification of a specific protein from the pool of cellular proteins, followed by Western blotting of a suspected interacting protein (1). Unlike immunofluorescence microscopy approaches, which provide correlative information about proteins sharing cellular space, co-IP can confirm direct often high-affinity physical interactions between two proteins or protein complexes of equal or better affinity. The validity of co-IP approaches can quickly deteriorate when forced expression of tagged proteins are used to investigate possible interactions, yet this approach, particularly when endogenous factors are assessed, can provide useful information about potential protein associations.

In recent years, novel approaches with improved sensitivity have been developed and applied. Two such examples are fluorescence resonance energy transfer (FRET) and proximity ligation assay (PLA). FRET is a biophysical method that can assess the presence of two adjacent proteins (usually within 10 nm of each other) by measuring energy transfer between two fluorophores. Each putative interacting protein is conjugated to one of the fluorophores and when coexpressed in a cell, interactions can be assessed by the generation of a predictable emission from the acceptor fluorophore as a result of proximal activation by the donor fluorophore (refs. 3–5; Fig. 1A). Numerous examples exist using FRET-based technologies. These include highly sensitive assessment of protein–protein interactions, such as that between Src kinase and matrix metalloproteinases in growth factor–mediated cancer cell growth (6), assessing the enzymatic activity of a complex, such as the kinase activity of BCR-ABL in chronic myeloid leukemia (CML; ref. 7), and in addition, the discovery of drugs that perturb protein–protein associations, such as those between estrogen receptor (ER) and its regulatory cofactors (8). FRET and derivations of FRET-based technologies allow live monitoring of protein interactions (9, 10) and have been powerful methodologies for investigating kinetic changes in protein associations due to the proximity required for signals to be recorded, but they require construction and expression of enzymatically enabled, modified version of the protein components of interest. This limits the scope to systems that are amenable to modification (such as cancer
cell lines) and uses structurally altered proteins, which may perturb normal physiologic function of the protein.

A distinct approach for assessing interactions between two candidate proteins is the PLA. This method uses probes against the two separate components, both of which are conjugated to oligonucleotides. Different substrates have been used, but a common theme involves coupling of the oligonucleotide sequence to antibodies (11), such that two antibodies against individual proteins will be juxtaposed by specific protein–protein interactions, enabling ligation of the oligonucleotide tails and detection of an association (reviewed in ref. 12; Fig. 1A). A major benefit of PLA is that it does not require modification or tagging of the proteins of interest, permitting assessment of endogenous interactions with great sensitivity. This was successfully employed to assess interactions between estrogen receptor (ER) and the signaling protein Src (13). Unexpectedly, an association between these two proteins from distinct pathways could be detected in both cancer cell line models and in primary breast cancer tumor material (13), permitting investigation of protein–protein interactions from clinical samples. However, the data suggested that 54% of tumors that were ER− by immunohistochemistry (IHC), contained ER−–Src interactions in the cytoplasm (13), which is in contrast to recent findings that only 1.5% of cancers that are ER+, as assessed by IHC, express any detectable cytoplasmic ER (14). In addition, it is well established that ER− patients with breast cancer gain no clinical benefit from drugs that inhibit ER function (15), a conclusion that is at odds with the finding that half of ER− tumors have detectable ER–Src associations, as measured by PLA (13). However, PLA offers exciting new diagnostic possibilities for measuring protein–protein interactions in clinical samples. Despite the opportunities afforded by

Figure 1. Examples of novel, sensitive directed approaches for assessing protein–protein interactions. A, FRET involves conjugating fluorophores to the two proteins investigated. If the two proteins are in close proximal space (within 10 nm), energy transfer occurs and emission can be measured. PLA involves attaching oligonucleotides to a substrate, in this case antibodies. Two antibodies against two potentially interacting proteins are combined and if the two proteins are associated with each other, the oligonucleotides can be ligated and detected by PCR-based methods. Both of these methods require prior knowledge of the putative interacting partner proteins. B, Indirect identification of associated proteins can be gleaned from analysis of transcription factor binding data, where enrichment of specific DNA motifs indicates potential associated proteins.
these types of methods, a fundamental limitation is that they require a prior knowledge of both proteins suspected to interact.

**Methods for Finding Unknown Associated Protein Networks**

A common requirement when studying proteins in cancer is the ability to find unknown proteins that associate with a factor of interest. This may be to assess the potential role of a protein in a disease of interest, such as cancer, to identify potential regulatory mechanisms of the protein, or to assess changes in protein associations in the presence of a treatment or a specific perturbation. To this end, numerous approaches exist, one of which is the Yeast two-hybrid approach, which uses the protein of interest as bait and identifies proteins that can interact with it (16).

The premise is that for functional transcriptional activity of a reporter gene to occur, two components need to be physically adjacent, one of which is a fixed core protein and the other is identified from a library of candidates. This traditionally exploits the GAL4 complex and the transcriptional readout is conducted in yeast cells and permits discovery of interacting proteins (17), although any interaction is assessed in a non-human context and is generally restricted to finding interactions between two proteins, rather than complexes of more than two factors.

An indirect, but powerful method for finding transcription factors that associate with a factor of interest is by identification of consensus binding motifs within global transcription factor binding events. Methods for mapping transcription factor binding sites, such as ChIP-seq, are relatively standard tools for many laboratories. The map of genome-wide contact points represent putative *cis*-regulatory elements, and they can be analyzed to identify DNA sequences that occur more often than expected (Fig. 1B).

This can reveal insight into other factors that cobind with the transcription factor that was mapped. ER was one of the first transcription factors to be mapped globally (18) and subsequent analysis of enriched motifs and biologic validation and characterization has revealed unexpected, but important ER-associated proteins. These include the identification of Forkhead motifs, revealing a role for the pioneer factor FoxA1 in mediating ER–chromatin interactions (19).

FoxA1 has subsequently been shown to be a critical determinant of ER function in cancer (20) and mammary gland development (21) and has been suggested as a putative therapeutic target for breast cancer (22). In addition, functional interplay between ER and AP2γ (23), PBX-1 (24) and GATA3 (25, 26) have resulted from the identification of DNA motifs for these specific factors within the global map of ER-binding sites.

Another method that is commonly used is purification of the protein of interest followed by mass spectrometry (MS) identification of associated proteins. One of the most popular methods is tandem affinity purification (TAP), which involves cloning tags onto your gene of interest resulting in expression of a fused protein (27). The protein-TAP tag fusion is expressed in a cell line and rounds of immunoprecipitation are conducted using the tags as a handle to isolate the protein complex, followed by gel electrophoresis to resolves interacting proteins and finally MS of bands present in the experimental lane but not the control IP lane (Fig. 2). The TAP-tagged method has been hugely successful in purifying unknown protein associations and has led to the identification of many novel cancer-related protein networks, including links between the BRCA1 complex and DNA damage components (28), an interplay between E2F and Myc pathways (29) and novel WNT/β-catenin–associated factors (ref. 30; Fig. 2). When used in yeast, the TAP tag is usually knocked into the endogenous locus of the gene of interest, replacing the wild-type gene (27, 31). Given difficulties with gene editing in mammalian systems, when applied to human systems, such as cancer cell lines, the TAP-tagged protein of interest is expressed on top of a background that may express the wild-type protein, resulting in competition between the purified tagged version and the endogenous untagged version. It is therefore common to use cell lines that do not express the protein of interest, circumventing the potential problems associated with competition between the endogenous protein and the tagged protein. This limits the physiologic relevance when looking at proteins with specific and limited tissue expression. As an example, the discovery of ER interacting proteins using TAP-tagged approaches was conducted in ER⁺ HeLa cells that were transduced with a tagged-ER expression construct (32). While this revealed interesting observations about novel ER-associated protein networks, including the role for NuRD-like repressors in specific cell-cycle phases, it identified protein associations that occur between exogenous ER and the proteins expressed in cervical cancer cell models (the HeLa cells used for the purification), rather than ER⁺ breast cancer cells that contain and depend on various components for a functional ER transcriptional pathway. One method for circumventing this problem is by expressing the tagged protein at the same time as a siRNA that is specific to the endogenous version, resulting in a reduction of the non-tagged endogenous protein and simultaneous compensation from the tagged version (31). While tagged purification approaches are common and have been successful for discovering interacting networks of proteins, there are potential negative effects that the presence of the tags may have on the protein structure of the bait and subsequently, the proteins that interact with the target factor. In addition, non-physiologic levels of the factor as a result of forced expression can result in artifactual interactions with chaperone proteins, due to the presence of excess, incorrectly folded protein (33). Furthermore, the amount of cells required can, in some cases, be extremely high. As an example, Nakatani and Ogryzko recommend using approximately 6 × 10⁹ cells for a relatively abundant protein complex of interest (34). This quantity of starting material is not uncommon (28, 35), but it restricts the applicability from primary tissue or clinical samples, where starting material is limited.
Methods for Identification of Endogenous Interactomes

The major technical benefit of using tagged gene approaches is the presence of a high affinity and robust purification step. However, there has been significant interest in developing and utilizing protein purification methods that permit isolation of endogenous proteins and the associated protein complex. This line of thought has benefited enormously by the rapid improvement of MS sensitivity, providing the opportunity to establish more physiologically relevant, but lower affinity purification methods. Such an approach would provide confidence that interactions are biologically and clinically relevant and are not artifacts of overexpressed proteins with tag sequences that could potentially alter the structure of the protein of interest. Ideally, an endogenous purification approach would be possible from limited starting material, enabling the possibility for discovery of unknown protein interactions from primary tumor material or in vivo cancer models.

Traditionally, experiments were conducted in parallel with control IPs or comparable matched experimental control conditions, so that interacting factors could be identified as enriched when compared to the control. However, the advent of methods for labeling whole pools of proteins, have enabled comparisons between samples with quantitative assurance. This has been achieved using multiple approaches, but a popular method that can be used in cell line models and whole animal experiments is SILAC (stable isotope labeling of amino acids in cell culture) labeling (36). SILAC labeling involves the addition of isotope containing amino acids to growing cells, so that the entire pool of proteins incorporate a detectable, introduced amino acid as a replacement for the normal unlabelled amino acid. This tends to involve isotopes of arginine or lysines (37). If different isotopes are incorporated into different populations, the protein pools following immunoprecipitation can be mixed and quantitative comparisons for each peptide can be computed, by comparing the ratio between different isotopes (37). Cancer cell lines can be grown in SILAC media, but interestingly, so can whole animals such as flies (38), worms (39), and even mice (40), permitting comparisons between experimental conditions, or between different tissues with quantitative confidence. SILAC labeling has been an invaluable tool when exploring global protein–protein purification methods. As an example, Matthias Mann’s lab developed a method called QUICK (quantitative immunoprecipitation combined with knockdown), which coupled SILAC labeling, siRNA knockdown of the target of interest, and an endogenous purification

![Figure 2. Two methods for global identification of associated proteins. A, the traditional tagging approach, where the protein of interest is expressed as a fusion with a tag. Large-scale cell culture is employed and the tag is purified by immunoprecipitation. The precipitated proteins are resolved by gel electrophoresis and enriched protein bands specific to the experimental condition are excised and identified by MS. B, QUICK involves SILAC labeling cells and transfecting either siRNA to the specific target or a control siRNA. Endogenous protein is pulled down and the two conditions are quantitative compared to identify proteins that are enriched and lost when the target is silenced.](image-url)
approach, using an antibody against the protein/s of interest (41). In this method, two populations of cells are labeled with different SILAC media, so that direct comparisons can be performed. In one population of cells, a siRNA is included against the factor of interest, resulting in silencing of the protein and generation of a matched control, where the only difference is the protein of interest (and potentially any downstream regulated targets of that protein). This was used to identify β-catenin–interacting proteins in a colon cancer cell line and Cbl-associated factors from 293T cells (41). Many non-specific interactions were detected but due to the presence of the matched control cell line, these could be discarded, resulting in the recovery of several known associated proteins. The total number of identified proteins was limited, with less than half a dozen interacting proteins identified for either factor. That said, the approach is rapid, requires fewer cells than most of the tagged approaches and uses endogenous proteins as a bait, an important distinction from other methods. Subsequent to this initial publication, QUICK has been used to identify Stat3 (42) and 14-3-3ζeta (43) interacting proteins, revealing numerous putative interacting proteins.

A separate endogenous purification approach has been described by the Qin and O’Malley labs. Their approach, similar to QUICK, uses endogenous purification with antibodies against the protein of interest, but it is not amenable for utilization of SILAC labeling. This method was initially used to assess specific proteins of interest (44), revealing insight into Mediator complexes, but was subsequently scaled up to more than 1,700 proteins (45). This requires significant starting material, with 20 liter cultures needed for each experimental condition from non-adherent cells and between 50 and 200 large tissue culture dishes for adherent cells (45). The approach taken in this method was to perform very mild, non-stringent wash conditions to retain transient or weak protein–protein interactions. The purified material is then separated by gel electrophoresis and lanes are cut into pieces, proteins are extracted, and samples are run on MS. This tour de force by the Qin and O’Malley labs, provides a useful resource for the scientific community, but is a technically demanding process that may not be immediately adaptable to non-specialist laboratories.

**RIME (Rapid Immunoprecipitation of Endogenous Proteins)**

Our laboratory has recently developed a derivation on these approaches, a method termed RIME (Rapid
Immunoprecipitation of endogenous proteins. One important addition to RIME is formaldehyde cross-linking, which helps maintain transient protein-protein interactions, permitting much more stringent wash conditions (Figure 3) (46). This results in decreased background, improved sensitivity and requires much less starting material: in the order of a couple of tissue culture dishes. In fact, as discussed below, RIME can be modified to assess endogenous protein-protein interactions in small pieces of primary tumor material. Cross-linking approaches for improved proteomic discovery of interactomes has been utilized successfully for many years. Although multiple cross-linking agents exist, formaldehyde is a commonly used method because it is cell permeable, allowing the capture of protein-protein interactions from living cells, it is reversible and it tends to cross-link within short distances (2Å; refs. 47, 48), meaning that only proteins that are adjacent or physically associated can be cross-linked together. Formaldehyde cross-linking has been used in combination with tagged proteins for identification of Myc interactors (47) and has even been used in animal studies to assess protein interactions (49). The primary benefit for laboratories such as ours, is that formaldehyde is used on a regular basis for transcription factor mapping approaches (Chromatin Immunoprecipitation: ChIP-seq), enabling capture and identification of potentially transient protein-DNA interactions. We therefore modified ChIP to allow identification of associated proteins, from an endogenous protein of interest. This involved formaldehyde cross-linking of cells, sonication of chromatin and purification using an antibody against the endogenous protein. We perform numerous stringent washes, decreasing the non-specific interactors, primarily to chromatographic supports, and therefore the background signal and we have introduced a modified tryptic digestion step that bypasses the need for gel electrophoresis and removes contamination from antibody IgG chains. This specific combination of steps permits purification of endogenous proteins from one or two tissue culture dishes and can be completed and ready for MS analysis in less than a day.

We applied RIME to identify ER interacting proteins from two tissue culture dishes, revealing 108 proteins that occurred in three independent RIME replicates, but in none of the 5 IgG control RIME experiments. The most enriched protein was ER itself, but also identified was many known ER interactors, including FoxA1, GATA3, p300, CBP, AIB1, NRP1, RXR, TLE1, AP2γ, and additional proteins that were validated by co-IP and ChIP-seq (46). When compared with the data for ER purification using the Qin and O’Malley approach (45), RIME for ER recovered substantially more proteins, including many classic and validated ER-interacting proteins (46). Furthermore RIME was easily adaptable to SILAC-labeled cells for quantitative comparisons between treatments or different experimental conditions (46), an important distinction from other approaches that require substantial starting material. As a further proof-of-principle, E2F4 was purified from growth arrested conditions and all known E2F4-interacting proteins, including other E2F proteins, all three pocket proteins (pRb, p107, and p130), and all three DP dimeric partners were identified (46). We have successfully utilized RIME for purification of more than 50 factors (Carroll lab, unpublished data), including experiments from primary mouse tissue and xenograft tumors. Interestingly, similar to the QUICK method (41), almost all successful RIME experiments are characterized by the presence of the protein target as the most enriched factor following MS analysis (46), constituting a useful tool for antibody validation before further proteomic or genomic experiments.

Proteins with poor quality antibodies that fail for ChIP-seq can be successfully purified by RIME (unpublished data from Carroll lab), implying that the robustness of the antibody for purification of interactomes is not linked to the success of the same antibody for ChIP-based methods. Given the ability to purify endogenous proteins from limited starting material, we coupled RIME with selective reaction monitoring, a sensitive but directed MS approach and attempted to perform protein-protein interaction assessment from clinical samples. We purified ER from 6 ER+ and one ER− primary breast tumor and in all 6 ER+ cancers (but not the control ER− tumor) we could identify multiple peptides for ER. In addition, we could find interactions between endogenous ER and some of the associated cofactors, including CBP and p300 in a subset of the breast cancers (46), but only in ER− cases. This suggests that protein-protein interactions can be detected in clinical tumor material, permitting the possibility of diagnostic approaches for assessing tumor outcome or potential drug responses, by monitoring differential protein interaction events that occur in one condition or the other. One example is a comparison between protein complexes associated with a clinically relevant phosphorylation event on a protein of interest versus the non-phosphorylated form.

Conclusions

Our ability to study cancer genomics on a global scale has become a reality. We can measure expression levels of all genes at once, silence all genes simultaneously, map all transcription factor binding sites in one experiment, and sequence the entire genome of a cancer relatively easily. However, our ability to quantitatively study comprehensive protein levels or protein interactions on a global scale has been slower to evolve. Having said that, the sensitivity of MS technologies has improved dramatically in recent years and methods are being developed for unbiased identification of unknown protein complexes. Importantly, these have involved purification of endogenous targets. One such method developed in our lab, RIME, is one of several proteomic methods for identifying protein-protein interactions. It benefits from being very quick, requires little starting material, is sensitive, and it works for endogenous proteins. For proteomics laboratories such as ours, RIME is easily adaptable and constitutes a useful tool in our toolbox. Regardless of the approach taken, incorporating information about endogenous protein interactions can be incredibly informative when studying pathways and regulatory mechanisms in cancer revealing unexpected insight into tumor biology.
Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: H. Mohammed
Writing, review, and/or revision of the manuscript: J.S. Carroll

Acknowledgments
The authors thank Clive D’Santos for reading the manuscript. The authors also thank the support of The University of Cambridge, Cancer Research UK, and Hutchison Whampoa Limited. The authors apologize to the numerous people whose work we were unable to cite.

Grant Support
Jason S. Carroll is supported by an ERC Starting Grant and an EMBO Young Investigator award.

Received August 25, 2013; accepted September 16, 2013; published OnlineFirst September 26, 2013.

References


Molecular Cancer Research

Approaches for Assessing and Discovering Protein Interactions in Cancer

Hisham Mohammed and Jason S. Carroll


Updated version
Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-13-0454

Cited articles
This article cites 49 articles, 18 of which you can access for free at:
http://mcr.aacrjournals.org/content/11/11/1295.full.html#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.