Review

Charting plant interactomes: possibilities and challenges

Mustafa Morsy[1], Satyanarayana Gouthu[1], Sandra Orchard[2], David Thorneycroft[2], Jeff F. Harper[1], Ron Mittler[1,3] and John C. Cushman[1]

[1] Department of Biochemistry and Molecular Biology, MS200, University of Nevada, Reno, NV 89557, USA
[2] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SD, UK
[3] Department of Plant Science, Hebrew University of Jerusalem, Jerusalem 91904, Israel

Protein–protein interactions are essential for nearly all cellular processes. Therefore, an important goal of post-genomic research for defining gene function and understanding the function of macromolecular complexes involves creating 'interactome' maps from empirical or inferred datasets. Systematic efforts to conduct high-throughput surveys of protein–protein interactions in plants are needed to chart the complex and dynamic interaction networks that occur throughout plant development. However, no single approach can build a complete map of the interactome. Here, we review the utility and potential of various experimental approaches for creating large-scale protein–protein interaction maps in plants. Bioinformatics approaches for curating and assessing the confidence of these datasets through inter-species comparisons will be crucial in achieving a complete understanding of protein interaction networks in plants.

Introduction: the importance of interacting

Protein-protein interactions (PPIs) are essential for almost all biological functions mediated by macromolecular machinery including replisomes, spliceosomes, ribosomes, proteosomes, signalosomes, catalytic enzyme modules, signal recognition complexes, and specialized protein complexes unique to plants such as light harvesting and photosystem complexes. Thus, identifying, quantifying, localizing, and modeling entire PPI networks (or 'interactomes') is a key prerequisite for understanding the biophysical basis of all cellular processes and for creating a framework to characterize the function of all proteins [1].

Plant proteomes of 30 000-40 000 proteins [2] are expected to have an estimated 75 000-150 000 interaction pairs, extrapolating from estimates of the size of the yeast interactome [3,4]. This illustrates the magnitude of the challenge for creating comprehensive interactome maps. Although several small- to mid-sized protein interaction studies have been published (see Supplementary Table 1), no large-scale experimental interactome mapping efforts have been reported in plants to date. The purpose of this review is to survey the frequently applied techniques for detecting and characterizing PPIs in plants and to highlight established and newly emerging high-throughput technologies that will become increasingly important to gaining a comprehensive, functional understanding of plant interactomes Table 1.

The challenge of studying protein–protein interactions

A wide range of approaches have been developed to characterize PPIs. These include both *in vitro* and *in vivo* experimental methods [5,6], and *in silico* approaches that are based on literature curation [7], evolutionarily conserved complexes [8–10], and network integration to infer functional modules (Figure 1) [1,11]. However, no single experimental approach can determine all PPIs, This is due, in part, to the diverse physiochemical properties, abundance, and subcellular localization patterns of individual proteins and the wide range of equilibrium dissociation constants of different protein complexes. Single approaches can supply only limited information about the key parameters of interactions or complexes, such as the oligomeric state of the interacting partner, the stoichiometric ratio within a complex, the relative affinity of interacting partners towards one another, the location of interaction sites, and the dependency of interactions on post-translational modifications [5]. Although comparative interactome approaches are useful among animal models, they provide an incomplete view of the plant interactome. This is because plants have evolved unique gene repertoires in order to accommodate their photoautotrophic and sessile lifestyle, and these functions are not present in other model organisms [12]. Therefore, it is crucial to generate plant-specific PPI interaction data using a variety of complementary approaches (Figure 1).

In vivo identification of protein–protein interactions

The ability to study PPIs within their native cellular context is crucial. *In vivo* approaches offer the ability to discover and study known or novel PPIs, often in real-time in living cells.

Yeast two-hybrid strategies

The yeast two-hybrid (Y2H) screening method provides a simple, robust binary readout of protein-protein interactions that has made it one of the most popular methods for the compilation of proteome-wide interactome maps [13]. The Y2H approach exploits the modular nature of transcription factors containing DNA-binding and activation domains. When these domains are split, the factor

Corresponding author: Cushman, J.C. (jcushman@unr.edu).

1360-1385/$ – see front matter doi:10.1016/j.tplants.2008.01.006 Available online 7 March 2008

Table 1. Summary of protein–protein interaction methodologies and their utility in low-, medium- and high-throughput studies

Methods	Throughput	Advantages	Disadvantages	Reference
***In vivo* approaches**				
Yeast two-hybrid (Y2H)	High	Mature; robust; widely used; can be automated	High false-positive rate, interactions limited to nucleus in classical system	[15,16]
Protoplast Y2H	Medium	Plant cell environment, potentially more sensitive than yeast system	High false-positive rate, interactions limited to nucleus; protoplasts technically challenging	[18]
Split-ubiquitin system (SUS)	Medium	Interactions can be detected in diverse cellular compartments	High false-positive rate; high background of enzymatic reporters	[20,22–24]
Bimolecular fluorescence complementation (BiFC)	Medium	Interactions can be detected in diverse cellular compartments	Multiple fusion orientations should be tested; non-specific interactions can be reported; background fluorescence	[26–29]
Split-luciferase system	Medium–high	Extremely low background	No subcellular localization information provided	[33–35]
Fluorescence resonance energy transfer (FRET)	Low	Quantitative; subcellular localization (time-resolved interactions possible)	Laborious; technically demanding; sophisticated imaging system required	[30]
Bioluminescence resonance energy transfer (BRET)	Low	Avoids energy donor photo-bleaching	Laborious; technically demanding; very sensitive charge-coupled device (CCD) camera required	[37,38]
***In vitro* approaches**				
Native chromatography or electrophoretic purification	Medium	Simple	Potentially poor resolution of complexes; high sample complexity	[39]
Immunopurification	Low	Reduced complexity sufficient for MS–MS identification	Specific antibody required, expensive; non-specific protein contamination	[41]
Single affinity purification-tagging	Medium–High	Efficient; flexible, low cost	Non-specific protein contamination	[42]
Tandem affinity purification (TAP) tagging	Medium–High	High sample purity; reduces non-specific proteins	Weak interactions can be disrupted; promiscuous protein interactors	[44,47–48]
Stable-isotope labeling of amino acids in cell culture (SILAC)	High	Uniform direct isotopic labeling without chemical differences	Incomplete (~80%) isotopic labeling; expensive; uses cell cultures	[53,54]
^{15}N-labeling	High	Inexpensive, complete labelling	Plants in liquid culture	[55,56]
Chemical crosslinking-MS using protein interaction reporter (PIRs)	High	Simple; large-scale identification of interactions possible	Reliable identification of interactions within complex protein mixtures problematic	[57,58]
Protein microrarrays	High	Simultaneous interrogation of large numbers of proteins	Technically challenging; requires large collections of expression ready cDNA clones	[61–63]
Surface plasmon resonance	Low–medium	No labeling required; kinetic studies possible	Sophisticated imaging system required	[64–66]
***In silico* approaches**				
Predicted interactomes	High	Experimental data from plants not essential	Limited by interolog annotations; incomplete	[74,85]

is functionally disabled. However, if each domain is fused to two interacting proteins as hybrid proteins (thus the name 'two-hybrid'), the function of the factor can be restored to transcribe one or more selectable marker or reporter genes that then select for the interaction event (Figure 2). Since its first introduction nearly 20 years ago [14], many improvements and enhancements have been developed, and these have been reviewed extensively [15,16].

Depending on the approach used, classical high-throughput Y2H assays has been estimated to reliably detect approximately 50% of true positives [13,17]. Despite its flaws, Y2H analysis has been used extensively in studies of plant binary protein interactions, and many novel protein interactions have been detected using this approach (Supplementary Table 1). These studies have provided researchers with invaluable entry points to explore many sub-networks of the *Arabidopsis* interactome. Nonetheless, interactions discovered or identified using Y2H analysis must be validated *in planta*.

The classical Y2H system has also been adapted for use in Arabidopsis protoplasts and has the potential of detecting protein interactions that are undetectable in heterologous systems [18]. Although this technique is suitable for the detection of nuclear protein interactions and can be adapted to a high-throughput format, protein interactions within other discrete subcellular compartments might be undetectable using this system. The production of high quality protoplasts can also make this methodology technically challenging [19]. Aside from the Y2H system, a variety of other options are now available to plant researchers to study PPIs in the context of living plant cells.

Protein (fragment) complementation strategies

Protein (fragment) complementation assays (PCAs) split an enzyme or a fluorescent reporter protein into two fragments that are fused to two interaction candidate proteins. Upon interaction, the split reporter protein is reconstituted (Figure 2). In contrast to classical Y2H strategies,

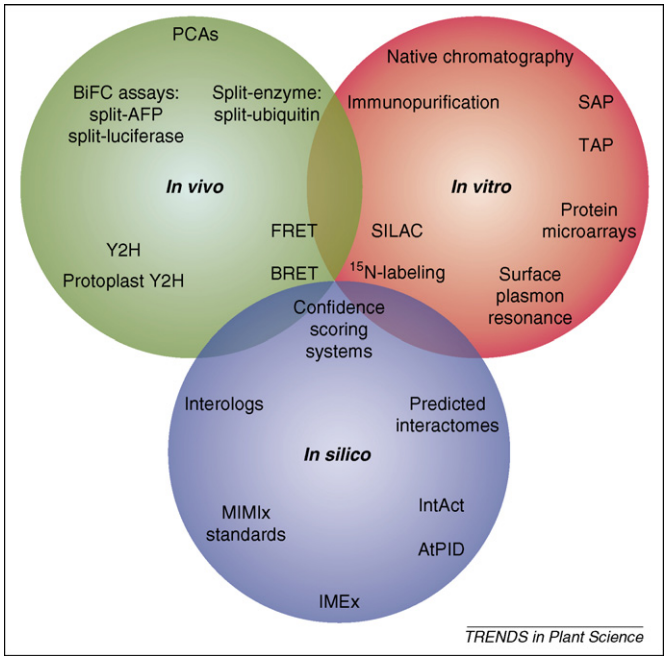

Figure 1. Conceptual overview of the major technological approaches used to survey PPIs *in vivo* and *in vitro,* and important concepts for *in silico* approaches used to curate and analyze interaction datasets. The overlap of the three circles illustrates the importance of integrating data from all three approaches to accurately define protein interactions. Abbreviations: AFP, auto-fluorescent proteins; AtPID, *A. thaliana* protein interaction databases; BiFC, bimolecular fluorescence complementation assays; BRET, bioluminescence resonance energy transfer; FRET, fluorescence resonance energy transfer; IMEx, international molecular exchange consortium for data deposition and sharing; MIMIx, minimum information required for reporting a molecular interaction experiment standards; PCAs, protein (fragment) complementation assays; SAP, single affinity purification; SILAC, stable-isotope labeling of amino acids in cell culture; TAP, tandem affinity purification; Y2H, yeast two hybrid.

PCAs permit protein interactions to be surveyed in diverse subcellular compartments [20]. Examples of this approach include the split-ubiquitin system (SUS) [21], which has been used in tomato [22] and *Arabidopsis* [23,24]. Since the initial use of ubiquitin, a wide variety of reporter enzymes have been used for split-protein-based screening systems [5,20,25].

To overcome limitations of high background signals and false positive rates associated with enzymatic reporters, bimolecular fluorescent complementation (BiFC) assays have been developed and used in plant cells [26,27]. BiFC is based on the formation of a fluorescent complex from two non-fluorescent fragments of auto-fluorescent proteins (AFPs) by the interaction between the proteins to which these fragments are fused (Figure 2). By using different combinations of AFP variants, numerous spectrally distinct complexes can form, and this enables multiplexed BiFC assays to be performed within a single living cell [28]. BiFC assays have been used successfully in a variety of plant species, including *Arabidopsis thaliana*, *Nicotiana* spp., and *Allium* sp. [27]. Versatile vector and delivery systems have been developed, enabling detection of specific PPIs in various subcellular compartments [29], and delivery of constructs by particle bombardment, agroinfiltration, or protoplast transfection [30].

Although the BiFC method is simple, rapid and suitable for medium-throughput screens, researchers should be aware of its potential drawbacks. Multiple fusion orientations should be tested, and the fusion proteins should be expressed at concentrations comparable to those of their endogenous counterparts [26,27]. Some very transient interactions can go undetected if the dissociation rate of a particular complex is faster than the rate of association between the fluorescent-protein fragments [31] or if the stability or molecular properties of the chimeric fusion proteins differ from those of the native proteins [26]. By contrast, if the intracellular concentrations of a split-AFP half is sufficiently high, then non-specific interactions might be reported, owing to the ability of some AFP fragments to form fluorescent chromophores [e.g. large N-terminal fragment of split-green fluorescent protein (GFP)] or complexes in the absence of fusions to specific interaction partners [e.g. yellow fluorescent protein (YFP)] [27].

An intrinsic drawback of the use of fluorescence-based detection systems in plants is the high degree of cellular auto-fluorescence. This can be partially overcome through the use of newly developed red fluorescent proteins (RFPs) that have long excitation and emission wavelengths and have been used successfully in BiFC systems [32]. Alternatively, PCAs that use split luciferase derived from either coral (*Renilla)* or firefly offer the ability to monitor PPIs using bioluminescence in living cells with extremely low background fluorescence [33]. In addition, they are useful for conducting large-scale PPI screens [34], as reported using Arabidopsis protoplasts in a 96-well plate format [35]. The disadvantage of this system is that it does not provide information about the subcellular location of the interaction. Although not a common occurrence, the possibility of 'bridging' proteins, which might mediate the interaction between two query proteins, should be considered when using PCA and BiFC approaches as screening assays, because these could result in apparent interactions that are not truly binary.

Fluorescence or bioluminescence resonance energy transfer

The availability of a wide variety of spectral variants of AFPs [36] has enabled the use of Förster (or fluorescence) resonance energy transfer (FRET) to detect PPIs in living plant cells (Figure 2). FRET is the radiation-less energy transition from a donor (e.g. GFP) to an acceptor fluorophore [e.g. red fluorescent protein (RFP)] that are in very close proximity (<10 nM) to one another and that have the appropriate steric orientation [30]. FRET and its variations, such as fluorescence life-time imaging microscopy (FLIM), are very powerful approaches for determining time-resolved, quantitative, subcellular localization of specific PPIs within single living cells. However, these techniques have not been used as widely as BiFC approaches, owing to their being more labor-intensive and technically demanding [30].

The limitations of high-background auto-fluorescence within plant cells, and the need for direct excitation of a fluorescence acceptor can be overcome by bioluminescence resonance energy transfer (BRET). Instead of using light emitted from an excited AFP, BRET employs a bioluminescence 'energy donor' such as blue-light emitting *Renilla* luciferase and either GFP or YFP as the 'energy acceptor'. BRET has been used successfully to monitor PPIs in

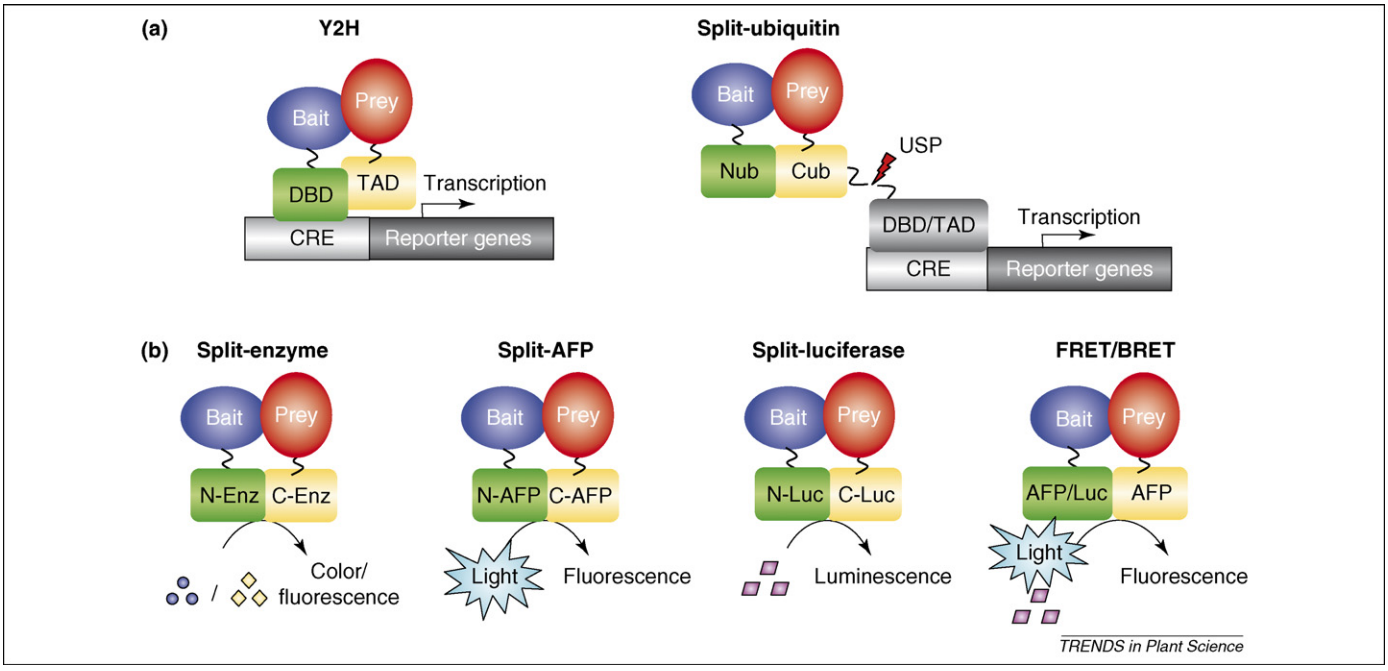

Figure 2. Methodologies for the detection of PPIs *in vivo*. **(a)** *In vivo* transcription-based reporter systems: Yeast two-hybrid (Y2H) detects PPI by way of one or more transcriptional reporters following reconstitution of the DNA-binding domain (DBD) and Trans-activation domain (TAD) of a split transcription factor. Split-ubiquitin detects PPI by release of a tethered transcription factor following activation of a ubiquitin-specific protease (USP) resulting from the reconstitution of the N-terminal half of ubiquitin (Nub) and the C-terminal half of ubiquitin (Cub); CRE, Cis-regulatory element. **(b)** *In vivo* protein complementation assays (PCAs): split-enzyme results in the generation of color from a chromogenic substrate (Blue circles) or fluorescence from a fluorescence substrate (yellow diamonds) upon reconstitution of the enzyme facilitated by bait-prey interaction; N-Enz, N-terminal half of enzyme. C-Enz, C-terminal half of enzyme. Split-AFP (auto-fluorescent protein) results in the generation of fluorescence following light excitation upon reconstitution of the AFP facilitated by bait-prey interaction; N-AFP, N-terminal half of AFP. C-AFP, C-terminal half of AFP. Split-luciferase results in the generation of fluorescence following excitation by a bioluminescence substrate (magenta parallograms) upon reconstitution of the luciferase enzyme facilitated by bait-prey interaction; N-Luc, N-terminal half of Luciferase. C-Luc, C-terminal half of Luciferase. Fluorescence resonance energy transfer (FRET) or bioluminescence resonance energy transfer (BRET) results in the generation of fluorescence following resonance energy transfer from emitted light from AFP or Luc from a bioluminescence substrate (magenta parallograms) upon close physical proximity of the AFP or Luc facilitated by bait-prey interaction. Adapted from [5].

Arabidopsis and tobacco cells [37,38]. BRET is more amenable to high-throughput screening than FRET because no internal excitation light source is needed and the energy donor does not become photo-bleached. However, light emission from BRET is generally rather dim, and detection of light emission by the fluorescence acceptor requires a very sensitive charge-coupled device (CCD) camera [38]. Although *in vivo* studies are often of primary importance for understanding PPI in the context of living cells, such approaches often require validation by complementary *in vitro* methodologies. *In vitro* approaches also offer very accessible and informative means of performing high-throughput PPI analyses.

In vitro identification of protein–protein interactions

The most direct approach towards elucidating PPIs *in vitro* is the preparative isolation and fractionation of native protein complexes using a variety of one-dimensional and multidimensional electrophoretic or chromatographic methods coupled with matrix-assisted laser desorption ionization-time-of-flight (MALDI–TOF) tandem mass spectrometry (MS–MS) [39]. Although this simple approach has been reported in only a small number of studies of plant protein complexes, it should be explored as a potentially useful complement to affinity-based purification strategies.

Affinity purification strategies

Affinity-based purification (AP) methods typically result in protein preparations of suitable purity and reduced complexity that they can be interrogated efficiently by MS–MS-based protein identification approaches [40]. Although direct AP approaches have been used in plants [41], these methods typically rely upon purification of a bait protein and interacting prey proteins using a custom antibody (preferably monoclonal) specific for the bait or a commercially available antibody to a peptide or protein epitope tag fused to the N or C termini of the bait (Figure 3). Epitope tagging strategies afford researchers greater flexibility and reduce costs associated with the production of antibodies to specific proteins [42]. The use of multiple affinity tags permits two consecutive or 'tandem' affinity purification (TAP) steps to be performed, usually under very mild and selective elution conditions [43]. Although the TAP approach reduces the amount of contaminating proteins, the increased processing involved can disrupt weaker interactions, lowering the range and abundance of proteins observed. TAP-tagging and MS–MS has been performed successfully in *Synechocystis sp.* PCC6803 [44], *A. thaliana* [44–46], rice [47], and to complement the Y2H approach in barley [48]. TAP-tagging systems with cleavage sites for a more specific and low-temperature active protease [49], and Gateway® compatible vectors have been developed to facilitate this approach in plants [50,51]. Using formaldehyde or glutaraldehyde stabilizes weak or transient protein interactions that might be lost during purification [42]. Epitope tagging approaches can be limited by various factors: the efficiency with which some fusion proteins compete with endogenous proteins for interaction with the bait [47]; the recurrent purification of artifactual

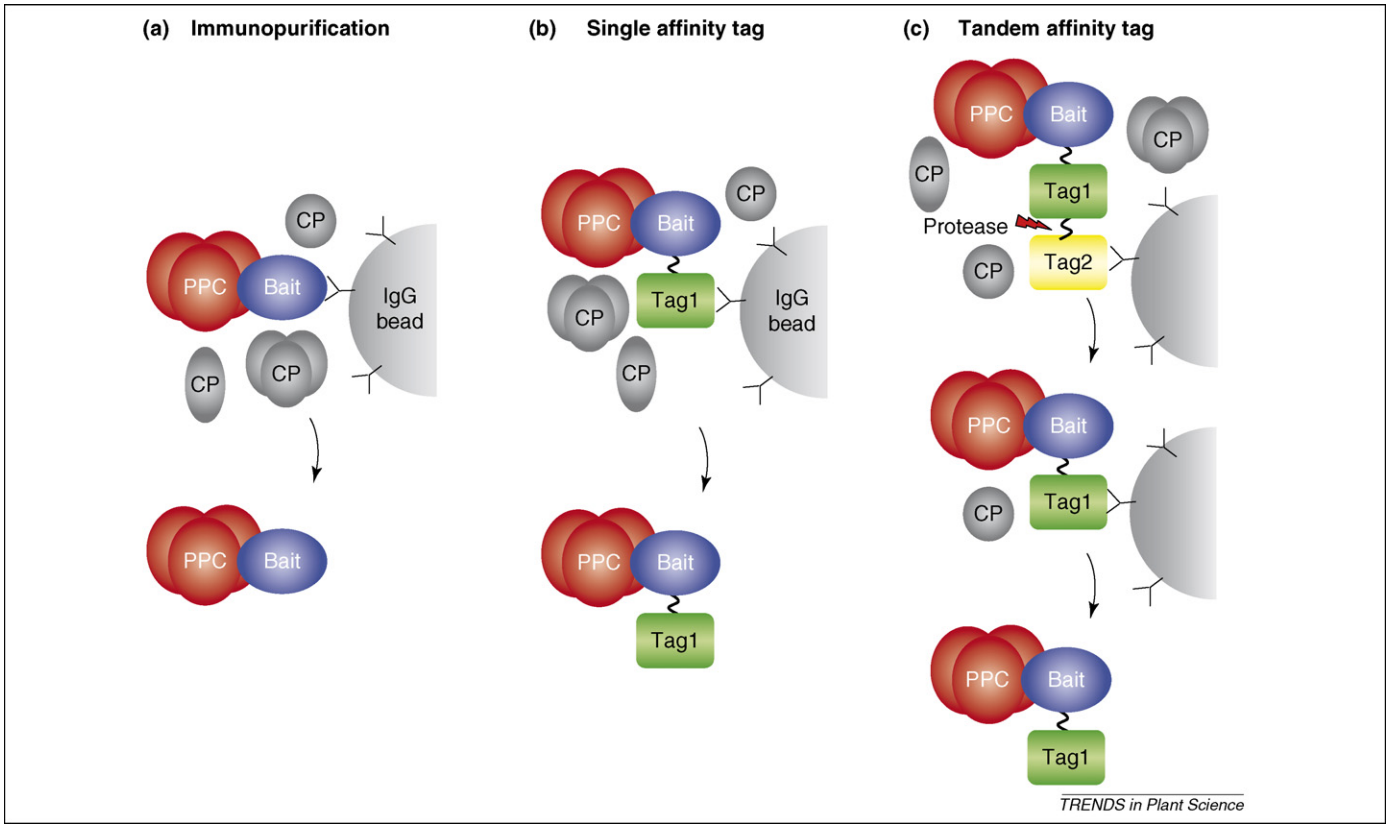

Figure 3. Methodologies for the detection of PPIs *in vitro*. *In vitro* affinity methods: **(a)** Immunoprecipitation with bead-immobilized antibody directed against a bait protein of interest; **(b)** single-affinity tag with bead-immobilized antibody directed against a universal protein tag (Tag1); **(c)** tandem affinity purification (TAP) tag using two sequential purification steps separated by a protease cleavage step with bead-immobilized antibodies directed against two universal protein tags (Tag1 and Tag2). Non-immunoaffinity tags can also be used. Abbreviations: CP, contaminating protein; PPC, prey protein complex. Adapted from [5].

proteins with promiscuous binding properties [47]; or N- or C-terminal fused tags disrupting PPIs or protein targeting of tagged proteins [49].

Quantitative MS strategies

An exciting development within the field of PPI studies has been the application of quantitative MS strategies, involving the incorporation of stable isotopes into proteins using metabolic labelling, and subsequent comparison of natural versus heavy isotopic ratios of the labeled proteins under control and test conditions [52].

Stable-isotope labeling of amino acids in cell culture (SILAC) involves growing cells in a suitable medium containing isotopically labeled amino acids [53]. Incorporation of isotopically labeled arginine into cell cultured proteomes and subsequent identification of 14-3-3 protein phosphorylation-specific interactions with bait peptides was demonstrated [54]. 14-3-3 proteins, named for their characteristic migration pattern following electrophoretic separation, participate in signaling by binding to a functionally diverse set of signaling proteins through recognition motifs that contain phosphorylated Ser or Thr residues [48]. Although plants can synthesis amino acids directly, cell cultures can be used in SILAC experiments, but incorporation of only 80% was achieved owing to the autotrophic nature of plant cells [54]. The advantage of SILAC-based methods is the efficient direct incorporation of the isotopic amino acid counterparts, which have no chemical differences, into the target proteome and the analysis of these in the same experiment. However, the technique is largely restricted to the use of cell cultures at present, primarily because of the high cost of isotopically labeled amino acids. These limitations can be overcome by the complete metabolic labeling of whole *Arabidopsis* seedlings in shaking liquid cultures using ^{15}N-enriched inorganic ammonium and nitrate salts as the sole nitrogen source. This approach appears to be a more robust and affordable strategy for isotopic labeling studies than SILAC [55,56] and should provide an effective tool for quantitative interactome studies in the future. However, one of the key advantages of SILAC is that peptides from labeled and unlabeled samples have the same mass and can be easily detected, whereas, in the N labeling approach, they have different masses and can be more difficult to identify accurately.

Interactome studies that rely on mass spectrometry-identifiable chemical cross-linkers called protein interaction reporters (PIRs) have enabled the large-scale identification of interacting proteins [57]. The main limitation to applying this approach on a proteome-wide scale is the complexity of the mass spectral information generated. However, recent simulations of PIR-type experiments have demonstrated the feasibility and constraints of the PIR strategy. Very complex protein mixtures may result in false discovery of PIRs due to random matching between peptides that coincidentally share similar masses [58].

Protein microarrays

Protein microarrays have great potential for high-throughput analysis of protein-protein interactions. Many different types of protein microarrays have been developed that

can be classified as either analytical, functional or reverse-phase [59,60]. Functional arrays suitable for assessing PPIs consist of full-length proteins or protein domains immobilized onto the surface of a glass slide. This permits the detection of protein interactions using fluorescence or chemiluminescent probes. A protein array displaying a collection of 1133 *A. thaliana* ORFs was fabricated and used to investigate the interactions of calmodulin (CaM) and CaM-like (CML) proteins [61]. Although multiple CaM or CML proteins bound to multiple partners, most interactions were specific to one or a few proteins, indicating that the CaM or CML network is more extensive than previously predicted [61].

Self-assembling or Nucleic Acid Programmable Protein Arrays (NAPPA) provide an alternative to protein microarrays that are fabricated by printing of proteins generated 'off-chip'. In these arrays, a collection of expression-ready cDNA clones [62] or unpurified polymerase chain reaction (PCR) products containing a protein expression cassette [63] is printed onto the array. The array is then bathed in a linked *in vitro* transcription-translation system to produce the protein array. Epitope tags fused to the expressed proteins enable them to remain immobilized on the surface of the array. With the development of large collections of expression-ready full-length cDNA clones for a growing number of plant species, the size of plant protein microarrays and their use can be expected to increase rapidly in the future.

Surface plasmon resonance

Surface plasmon resonance (SPR) exploits changes that occur in the local refractive index at a metal surface when immobilized target molecules interact with analyte containing potential interactors [64]. Proteins are unlabelled, which avoids many processing steps associated with other approaches. The technique also enables kinetic studies to be carried out, providing valuable information about the binding affinities, and association and dissociation kinetics of PPIs. Recent improvements in SPR imaging technologies have significantly improved the potential utility of this approach for studying PPIs. Such improvements include the use of charge-coupled device (CCD) cameras to monitor light scattering [65], high-throughput array formats [64], and coupling of arrays with MALDI-TOF MS analysis [66]. Although SPR has not been widely applied to investigate plant PPIs, these recent technical advances will make this approach increasingly useful for plant scientists.

Despite the utility of experimental methods in discovering novel PPIs or in the validation of known PPIs, all are prone to experimental error. Therefore, it is imperative that PPI datasets be validated rigorously by one or more complementary experimental techniques or that the quality of the reported interactions is evaluated using *in silico* approaches that can predict or assess the relative validity of the PPI in question.

In silico **analysis of protein interactions**

Various confidence-scoring systems have been developed to assess the reliability of PPI datasets derived from high-throughput approaches. These systems are based on one or more of the following metrics: co-expression, co-localisation, co-evolution, functional similarity, occurrence of orthologs or interologs (i.e. interactions conserved across species), number of observations of the interaction, type of interaction (e.g. direct, physical association, genetic interaction, etc.), topological characteristics of protein networks across species, or reciprocity of protein interactions [1,11,67–69].

Comparison and assessment of these different approaches has been difficult because many of them apply different combinations of confidence assignment metrics or false-positive identification rates. Furthermore, they are confounded by the lack of experimental validation of 'untested' or 'tested but not proven interactions' [67]. Another major caveat of *in silico* analyses is that protein function is frequently assigned using gene ontology (GO) terms, which are widely accepted, controlled vocabulary terms defining gene and gene product attributes [70]. In the case of *Arabidopsis*, the majority of annotations are inferred from sequence or structural similarity, and not from direct experimental evidence. Thus, the reliability of these assignments is questionable. Approximately 45% of the proteins in *Arabidopsis* are annotated by the term 'unknown' [71], indicating that coverage and depth of GO annotation in *Arabidopsis* is far from complete. In addition, some proteins have multiple functions [72,73], which can complicate GO assignments. Several recent studies have used homology-based predictive approaches to determine possible interaction networks in *Arabidopsis*.

Predicted interactomes

Predicted interologs identification has been used as a predictor of protein interaction on the premise that orthologous proteins, which are known to interact in one organism, can interact in the organism under study [8–10]. This approach was used recently to predict the interactome of *Arabidopsis* [74]. The study used confidence scoring based on the number of observations and experimental approaches that detected the interaction, together with subcellular location and co-expression data, to indicate the reliability of the interaction. For the 3617 conserved *Arabidopsis* proteins identified in the study, a total of 1159 'high confidence', 5913 'medium confidence' and 12 907 'low confidence' interactions were determined. A related study reported the prediction of 23 396 interactions in *Arabidopsis*, using several approaches, including interolog identification. These data are available through the *A. thaliana* protein interaction database (AtPID, http://atpid.biosino.org/) [75].

Predicted interactomes of this type can be useful in identifying networks of proteins that are highly conserved among species and for suggesting possible roles for unknown proteins [69]. However, the interolog approach is limited by the number of predicted orthologs found for the organism studied. In *Arabidopsis*, approximately 10 000 orthologs are detected in at least one reference species (i.e. in yeast, *Caenorhabditis elegans*, *Drosophila* or human). This amounts to less than half of the *Arabidopsis* proteome. Furthermore, the assignment of interologs is based on global protein sequence similarity, which could be insensitive to residue-specific requirements for interaction

specificity and affinity [76]. Genome duplication events have also been a pervasive force in plant evolution. These phenomena have resulted in expanded and plant-specific gene families involved in transcriptional regulation, signal transduction, and development [2]. Gene families are often composed of members differing from each other by a few amino acid residues. This close similarity can make unambiguous assignment of interologs problematic and can result in a contrived view of biological complexity. In addition, highly disordered proteins or proteins that lack recognized secondary structural motifs can play important roles in PPIs. However, low sequence similarity within disordered regions imply that homology-based approaches are unsuitable for the detection of functionally equivalent disordered proteins across species [77]. Predictive studies are still at an early stage. However, they are expected both to improve rapidly as more interactome data become available in plants and to provide powerful tools for generating hypotheses for future experimental investigations [67].

Databases and standards

Several protein interaction databases, including IntAct [78], Molecular Interaction Database (MINT) [79], AtPID [75], Database of Interacting Proteins (DIP) [80], Biomolecular Interaction Network Database (BIND) [81], and Biological General Repository for Interaction Datasets (BioGRID) [82], have emerged to organize, store and make PPI data available for analysis by the research community. Two databases are notable for their focus on plants: IntAct, which actively curates all plant protein interaction data from the literature and submitted datasets; and AtPID, which contains curated and predicted interaction data. The *Arabidopsis* Interactions Viewer (http://bar.utoronto.ca/interactions/cgi-bin/*Arabidopsis*_interactions_viewer.cgi) permits the visualization of selected interaction datasets from Geisler-Lee *et al.* [74] and Popescu *et al.* [61]. Readers should be cautioned that manually curated data are not error-free, owing to inconsistent curation standards and human error. However, manually curated data appear to be of higher quality than those obtained by automated approaches such as text-mining. Improvement in data curation is an ongoing process involving cooperation between databases, and the continued development of curation standards.

The work of the Human Proteome Organization Proteomics Standards Initiative (HUPO–PSI; http://www.pside-v.info/) has greatly improved the ability to combine or compare interaction data from diverse sources. Currently, all the major public domain databases export their data in a common standardized, data descriptive Extensible Markup Language (XML) interchange format annotated by a series of controlled vocabularies [83]. The data can be accessed from the IntAct molecular interaction database [78] in both XML and text-file formats.

In addition, HUPO–PSI has published guidelines to outline the Minimum Information required for reporting a Molecular Interaction Experiment (MIMIx) [83]. MIMIx outlines the minimum information required to describe all relevant aspects of the interaction experiment while minimizing the burden placed on the scientists generating the data. The establishment of the International Molecular Exchange Consortium (IMEx) encourages data deposition and sharing by all of the participating major public domain databases, the use of standardized protein identifiers, and clear experimental descriptors of the techniques used, and it ensures maximal data availability to the scientific community (http://imex.sf.net) [84].

Conclusions

We have surveyed the major methodologies for characterizing PPIs *in planta*, *in vitro* and *in silico*. As high-throughput technologies are applied to plant systems, we can expect rapid progress towards a comprehensive examination of plant interactomes, especially within selected plant models such as *Arabidopsis* and rice, for which large ORF collections of full-length cDNAs are already available. Current models of predicted plant interactomes will become increasingly robust and a better representative of the dynamic nature of the interactome by a steady accumulation of interaction data from a variety of experimental approaches. An important application of these data will be to inform plant scientists about the function of unknown proteins by virtue of their relative context within macromolecular complexes. Such novel information will greatly improve our understanding of the mechanisms that control protein interaction and organize molecular structures in plants.

Note added in proof

A firefly luciferase PCA was described recently by Chen *et al.* (2008) for the detection of PPIs in transient assays using *Arabidopsis* protoplasts or Agrobacterium-mediated transient expression in *N. benthamiana*. See: Chen, H. *et al.* (2008) Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiology* 146, 368–376.

Acknowledgements

Supported by the National Science Foundation (NSF-0420033; NSF-0420152) and the University of Nevada Agricultural Experiment Station, publication number 03077064. This publication was made possible by National Institutes of Health (NIH) grant number P20 RR-016464 from the IDeA Network of Biomedical Research Excellence (INBRE) Program of the National Center for Research Resources (NCRR) through its support of the Nevada Genomics Center and the Nevada Center for Bioinformatics. We acknowledge the contributions of all of the members of the IntAct research team, including Yasmin Alam-Faruque, Bruno Aranda, Alan Bridge, Cathy Derow, Emily Dimmer, Micheal Feuermann, Rachel Huntley, Jyoti Khadake, Sam Kerrien, Catherine Leroy, Luisa Montecchi-Palazzi, Sandra Orchard, David Thorneycroft, Rolf Apweiler and Henning Hermjakob.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tplants.2008.01.006.

References

1 Sharan, R. and Ideker, T. (2006) Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* 24, 427–433
2 Sterck, L. *et al.* (2007) How many genes are there in plants (. . . and why are they there)? *Curr. Opin. Plant Biol.* 10, 199–203
3 Grigoriev, A. (2003) On the number of protein–protein interactions in the yeast proteome. *Nucleic Acids Res.* 31, 4157–4161

4 Bader, G.D. and Hogue, C. (2002) Analyzing yeast protein–protein interaction data obtained from different sources. *Nat. Biotechnol.* 20, 991–997
5 Piehler, J. (2005) New methodologies for measuring protein interactions *in vivo* and *in vitro*. *Curr. Opin. Struct. Biol.* 15, 4–14
6 Meng, J.J. *et al.* (2005) Methods to study protein–protein interactions. *Methods Mol. Biol.* 289, 341–358
7 Reguly, T. *et al.* (2006) Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J. Biol.* 5, 11
8 Matthews, L.R. *et al.* (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or "interologs". *Genome Res.* 11, 2120–2126
9 Yu, H. *et al.* (2004) Annotation transfer between genomes: protein–protein interlogs and protein-DNA regulogs. *Genome Res.* 14, 1107–1118
10 Sharan, R. *et al.* (2005) Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1974–1979
11 Sharan, R. *et al.* (2007) Network-based prediction of protein function. *Molecular Systems Biology* 3, 88 DOI: 10.1038/msb4100129
12 Gollery, M. *et al.* (2006) What makes species unique? The contribution of proteins with obscure features. *Genome Biol.* 7, R57 DOI: 10.1186/gb-2006-7-7-r57
13 Parrish, J.R. *et al.* (2006) Yeast two-hybrid contributions to interactome mapping. *Curr. Opin. Biotechnol.* 17, 387–393
14 Fields, S. and Song, O. (1989) A novel genetic system to detect protein–protein interactions. *Nature* 340, 245–246
15 Causier, B. and Davies, B. (2002) Analysing protein–protein interactions with the yeast two-hybrid system. *Plant Mol. Biol.* 50, 855–870
16 Causier, B. (2004) Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom. Rev.* 23, 350–367
17 von Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403
18 Ehlert, A. *et al.* (2006) Two-hybrid protein–protein interaction analysis in *Arabidopsis* protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J.* 46, 890–900
19 Yoo, S. *et al.* (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572
20 Michnick, S.W. (2003) Protein fragment complementation strategies for biochemical network mapping. *Curr. Opin. Biotechnol.* 14, 610–617
21 Stagljar, I. and Fields, S. (2002) Analysis of membrane protein interactions using yeast-based technologies. *Trends Biochem. Sci.* 27, 559–563
22 Reinders, A. *et al.* (2002) Protein–protein interactions between sucrose transporters of different affinities colocalized in the same enucleate sieve element. *Plant Cell* 14, 1567–1577
23 Schulze, W.X. *et al.* (2003) Interactions between co-expressed *Arabidopsis* sucrose transporters in the split-ubiquitin system. *BMC Biochem.* 4, 3
24 Obrdlik, P. *et al.* (2004) K^+ channel interactions detected by a genetic system optimized for systematic studies of membrane protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12242–12247
25 Remy, I. and Michnick, S. (2007) Application of protein–fragment complementation assays in cell biology. *Biotechniques* 42, 137–141
26 Ohad, N. *et al.* (2007) The analysis of protein–protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiol.* 145, 1090–1099
27 Kerppola, T.K. (2008) Bimolecular fluorescence complementation: visualization of molecular interactions in living cells. *Methods Cell Biol.* 85, 431–470
28 Hu, C.D. and Kerppola, T. (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* 21, 539–545
29 Citovsky, V. *et al.* (2006) Subcellular localization of interacting proteins by bimolecular fluorescence complementation in planta. *J. Mol. Biol.* 362, 1120–1131
30 Bhat, R. *et al.* (2006) The visible touch: *in planta* visualization of protein–protein interactions by fluorophore-based methods. *Plant Methods* 2, 12 DOI: 10.1186/1746-4811-2-12 (http://www.plantmethods.com)
31 Kerppola, T.K. (2006) Visualization of molecular interactions by fluorescence complementation. *Nat. Rev. Mol. Cell Biol.* 7, 449–456
32 Fan, J.Y. *et al.* (2007) Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein–protein interactions in living cells. *Biochem. Biophys. Res. Commun.* 367, 47–53
33 Paulmurugan, R. and Gambhir, S. (2005) Firefly luciferase enzyme fragment complementation for imaging in cells and living animals. *Anal. Chem.* 77, 1295–1302
34 Paulmurugan, R. and Gambhir, S. (2007) Combinatorial library screening for developing an improved split-firefly luciferase fragment-assisted complementation system for studying protein–protein interactions. *Anal. Chem.* 79, 2346–2353
35 Fujikawa, Y. and Kato, N. (2007) Split luciferase complementation assay to study protein–protein interactions in *Arabidopsis* protoplasts. *Plant J.* 52, 185–195
36 Müller-Taubenberger, A. and Anderson, K.I. (2007) Recent advances using green and red fluorescent protein variants. *Appl. Microbiol. Biotechnol.* 77, 1–12
37 Subramanian, C. *et al.* (2006) A suite of tools and application notes for *in vivo* protein interaction assays using bioluminescence resonance energy transfer (BRET). *Plant J.* 48, 138–152
38 Xu, X. *et al.* (2007) Imaging protein interactions with bioluminescence resonance energy transfer (BRET) in plant and mammalian cells and tissues. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10264–10269
39 Liu, X. *et al.* (2008) Toward chromatographic analysis of interacting protein networks. *J. Chromatogr. A.* 1178, 24–32
40 Köcher, T. and Superti-Furga, G. (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* 4, 807–815
41 Toueille, M. *et al.* (2007) The elongation factor 1A: a novel regulator in the DNA replication/repair protein network in wheat cells? *Plant Physiol. Biochem.* 45, 113–118
42 Berggard, T. *et al.* (2007) Methods for the detection and analysis of protein–protein interactions. *Proteomics* 7, 2833–2842
43 Rigaut, G. *et al.* (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032
44 Brown, A.P. *et al.* (2006) Tandem affinity purification tagging of fatty acid biosynthetic enzymes in *Synechocystis* sp. PCC6803 and *Arabidopsis thaliana*. *J. Exp. Bot.* 57, 1563–1571
45 Rohila, J.S. *et al.* (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* 38, 172–181
46 Van Leene, J. *et al.* (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 6, 1226–1238
47 Rohila, J.S. *et al.* (2006) Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* 46, 1–13
48 Schoonheim, P.J. *et al.* (2007) A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. *Plant Physiol.* 143, 670–683
49 Rubio, V. *et al.* (2005) An alternative tandem affinity purification strategy applied to *Arabidopsis* protein complex isolation. *Plant J.* 41, 767–778
50 Nakagawa, T. *et al.* (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* 104, 34–41
51 Nakagawa, T. *et al.* (2007) Improved gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* 71, 2095–2100
52 Bantscheff, M. *et al.* (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031
53 Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958
54 Gruhler, A. *et al.* (2005) Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol. Cell. Proteomics* 4, 1697–1709
55 Huttlin, E.L. *et al.* (2007) Comparison of full versus partial metabolic labeling for quantitative proteomics analysis in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 6, 860–881
56 Nelson, C.J. *et al.* (2007) Implications of 15N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. *Proteomics* 7, 1279–1292
57 Tang, X. *et al.* (2005) Mass spectrometry identifiable cross-linking strategy for studying protein–protein interactions. *Anal. Chem.* 77, 311–318

58 Anderson, G.A. *et al.* (2007) Informatics strategies for large-scale novel cross-linking analysis. *J. Proteome Res.* 6, 3412–3421
59 LaBaer, J. and Ramachandran, N. (2005) Protein microarrays as tools for functional proteomics. *Curr. Opin. Chem. Biol.* 9, 14–19
60 Hall, D.A. *et al.* (2007) Protein microarray technology. *Mech. Ageing Dev.* 128, 161–167
61 Popescu, S.C. *et al.* (2007) Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4730–4735
62 Ramachandran, N. *et al.* (2004) Self-assembling protein microarrays. *Science* 305, 86–90
63 Angenendt, P. *et al.* (2006) Generation of high density protein microarrays by cell-free *in situ* expression of unpurified PCR products. *Mol. Cell. Proteomics* 5, 1658–1666
64 Campbell, C.T. and Kim, G. (2007) SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. *Biomaterials* 28, 2380–2392
65 Savchenko, A. *et al.* (2007) Imaging technique for the screening of protein–protein interactions using scattered light under surface plasmon resonance conditions. *Anal. Chem.* 79, 1349–1355
66 Nedelkov, D. (2007) Development of surface plasmon resonance mass spectrometry array platform. *Anal. Chem.* 79, 5987–5990
67 Suthram, S. *et al.* (2006) A direct comparison of protein interaction confidence assignment schemes. *BMC Bioinformatics* 7, 360
68 Chen, J. *et al.* (2006) Detecting functional modules in the yeast protein–protein interaction network. *Bioinformatics* 22, 2283–2290
69 Kiemer, L. and Cesareni, G. (2007) Comparative interactomics: comparing apples and pears? *Trends Biotechnol.* 25, 448–454
70 Zhou, X. and Su, Z. (2007) EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. *BMC Genomics* 8, 246
71 Haas, B.J. *et al.* (2005) Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release. *BMC Biol.* 3, 7
72 Jeffery, J.C. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* 19, 415–417
73 Cho, Y-H. *et al.* (2006) Regulatory functions of nuclear hexokinase1 complex in glucose signaling. *Cell* 127, 579–589
74 Geisler-Lee, J. *et al.* (2007) A predicted interactome for *Arabidopsis thaliana*. *Plant Physiology*
75 Cui, J. *et al.* (2008) AtPID: *Arabidopsis thaliana* protein interactome database an integrative platform for plant systems biology. *Nucleic Acids Res.* 36, D999–D1008
76 Uhrig, J.F. (2006) Protein interaction networks in plants. *Planta* 224, 771–781
77 Tompa, P. *et al.* (2005) Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* 30, 484–489
78 Kerrien, S. *et al.* (2007) IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.* 35, D561–D565
79 Chatr-aryamontri, A. *et al.* (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.* 35, D572–D574
80 Salwinski, L. *et al.* (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451
81 Gilbert, D. (2005) Biomolecular interaction network database. *Brief. Bioinform.* 6, 194–198
82 Stark, C. *et al.* (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539
83 Kerrien, S. *et al.* (2007) Broadening the horizon - level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.* 5, 44
84 Orchard, S. *et al.* (2007) Submit Your Interaction Data the IMEx Way: a Step by Step Guide to Trouble-free Deposition. *Proteomics* 7, 28–34
85 Cui, J. *et al.* (2008) AtPID: *Arabidopsis thaliana* protein interactome database—an integrative platform for plant systems biology. *Nucleic Acids Res.* 36, D999–D1008