Plant Protein Interactomes

Pascal Braun,[1] Sébastien Aubourg,[2] Jelle Van Leene,[3,4] Geert De Jaeger,[3,4] and Claire Lurin[2]

[1]Department of Plant Systems Biology, Center for Life and Food Sciences Weihenstephan, Technische Universität München (TUM), 85354 Freising-Weihenstephan, Germany; email: pbraun@wzw.tum.de

[2]Unité de Recherche en Génomique Végétale (URGV), UMR INRA/UEVE 1165–ERL CNRS 8196, 91057 Evry, France; email: aubourg@evry.inra.fr, lurin@evry.inra.fr

[3]Department of Plant Systems Biology, Vlaams Instituut voor Biotechnologie (VIB), 9052 Ghent, Belgium

[4]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium; email: jelee@psb.vib-ugent.be, gejae@psb.vib-ugent.be

Annu. Rev. Plant Biol. 2013. 64:161–87

First published online as a Review in Advance on January 16, 2013

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

This article's doi: 10.1146/annurev-arplant-050312-120140

Keywords

protein-protein interactions, functional genomics, networks, model plants and crops

Abstract

Protein-protein interactions are a critical element of biological systems, and the analysis of interaction partners can provide valuable hints about unknown functions of a protein. In recent years, several large-scale protein interaction studies have begun to unravel the complex networks through which plant proteins exert their functions. Two major classes of experimental approaches are used for protein interaction mapping: analysis of direct interactions using binary methods such as yeast two-hybrid or split ubiquitin, and analysis of protein complexes through affinity purification followed by mass spectrometry. In addition, bioinformatics predictions can suggest interactions that have evaded detection by other methods or those of proteins that have not been investigated. Here we review the major approaches to construct, analyze, use, and carry out quality control on plant protein interactome networks. We present experimental and computational approaches for large-scale mapping, methods for validation or smaller-scale functional studies, important bioinformatics resources, and findings from recently published large-scale plant interactome network maps.

Contents

INTRODUCTION

Despite the increasing number of fully sequenced higher-plant genomes and the technical progress in experimental genomics, many genes remain functionally uncharacterized, and the understanding of others is often incomplete. To accelerate functional studies, numerous high-throughput approaches have been developed in the past 10 years. Since the completion of the *Arabidopsis* genome sequence in 2000, transcriptomics and reverse genetics in particular have accelerated the functional characterization of its thousands of proteins. However, because proteins usually function in macromolecular complexes, it is difficult (or even impossible) to gain an understanding of a given protein's function without knowing which other proteins it physically and functionally interacts with. Conversely, interactions of an unknown protein with proteins of well-characterized function can be an important clue about the function of the former (9, 10).

For these reasons, there is an increasing interest in characterizing protein-protein interaction (PPI) networks, also referred to as interactome networks (see also sidebar, Functionally Annotating the *Arabidopsis* Genome by Revealing the Protein Interactome, and **Figure 1**). Most approaches for mapping these networks require cloned protein-encoding sequences—i.e., open reading frames (ORFs)—and are commonly referred to as clone-based proteomics.

In this review, we summarize the most common methods of characterizing plant interactomes and compare the advantages and caveats associated with each. We present both experimental and computational systematic approaches as well as targeted approaches and review the available bioinformatics tools for interaction prediction, network visualization, and data analysis.

PPI: protein-protein interaction

Open reading frame (ORF): full-length coding sequence without the 5′ and 3′ regulatory sequences of the cDNA

METHODS FOR MAPPING PLANT PROTEIN INTERACTIONS

Unlike transcriptomics and genetics approaches, most interactomics methods require the expression of fusion proteins using cloned ORFs. To facilitate such studies, high-throughput approaches based on recombinational cloning have been developed that allow the cloning of hundreds of proteins simultaneously (13, 42, 111, 151). Several ORF collections, some with a particular biological focus, have been constructed for *Arabidopsis* (e.g., 151, 156). In total, ORFs for almost 15,000 loci are now available as Gateway® entry clones from the *Arabidopsis* Biological Resource Center (95). Being expensive and time consuming to build, ORF library construction is generally restricted to model organisms and has not been generalized to many plants. However, some ORF libraries are available for rice (**http://www.grassius.org/tfomecollection.html**) and for *Brachypodium*

(15). The availability of ORF libraries has eliminated an important bottleneck of interactome mapping and allows researchers to focus on the actual interaction mapping experiments.

Untargeted and Systematic Methods

The unbiased, systematic investigation of interactions among a set of proteins provides the opportunity to uncover features of protein connectivity that can lead to insights beyond the discovery of novel biochemical relationships between two proteins of interest (9, 147). For example, a protein may be found to be central to a signaling network and function with different low-connectivity "specificity" proteins to regulate different processes (43). For the systematic investigation of global interactome networks in particular, the employed mapping methods must meet strict requirements regarding precision, cost, and throughput.

Heterologous expression in yeast. Yeast has several features that make it an attractive organism for high-throughput studies, particularly in interaction mapping. These include low operating costs, low amounts of DNA needed for transformation, and the existence of mating, which facilitates the introduction of two plasmids into a diploid yeast cell. The most widely used assays are the yeast two-hybrid (Y2H) (14, 29, 152) and split-ubiquitin (sUbq) (103) systems, although others exist (88).

Yeast two-hybrid assay. Y2H (33) has been widely used in animal and plant interaction mapping experiments on the scale of individual proteins, biological modules, protein families, and whole proteomes (2, 17, 51, 63, 74, 119, 139, 158) (**Figure 2*a***, top). Several Y2H versions exist that differ in the respective protein fragments used as activation and DNA-binding domains, vectors, and strains. It is important to recognize that such differences influence whether an interaction is detected (9, 11, 21) owing to the underlying inherent differences in the fusion protein geometry, expression levels, and cell physiology. For systematic interactome studies in particular, the specificity and sensitivity of any assay implementation should be determined using an adequately sized positive and random reference set (PRS and RRS, respectively) (9, 29). (For a more detailed discussion of binary interaction

FUNCTIONALLY ANNOTATING THE *ARABIDOPSIS* GENOME BY REVEALING THE PROTEIN INTERACTOME

The *Arabidopsis* proteome can be represented as an iceberg (**Figure 1**): For approximately 90% of *Arabidopsis* genes (the hidden part of the iceberg) we have no clues to the physiological function of the gene product, although for approximately two-thirds of these we can glean some idea of the molecular function from sequence similarity. In contrast, detailed experimental data allowing precise functional characterization are available for only a tiny fraction of the genes (the tip of the iceberg). Any protein-protein interaction between the tip and the hidden part will help in understanding the functions of unknown proteins. Furthermore, any interaction between plant proteins and symbiotic or pathogenic organism proteins will also increase understanding of the plant proteome.

Y2H: yeast two-hybrid

sUbq: split ubiquitin

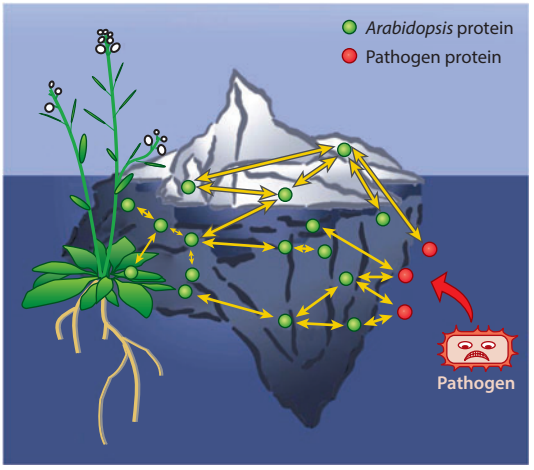
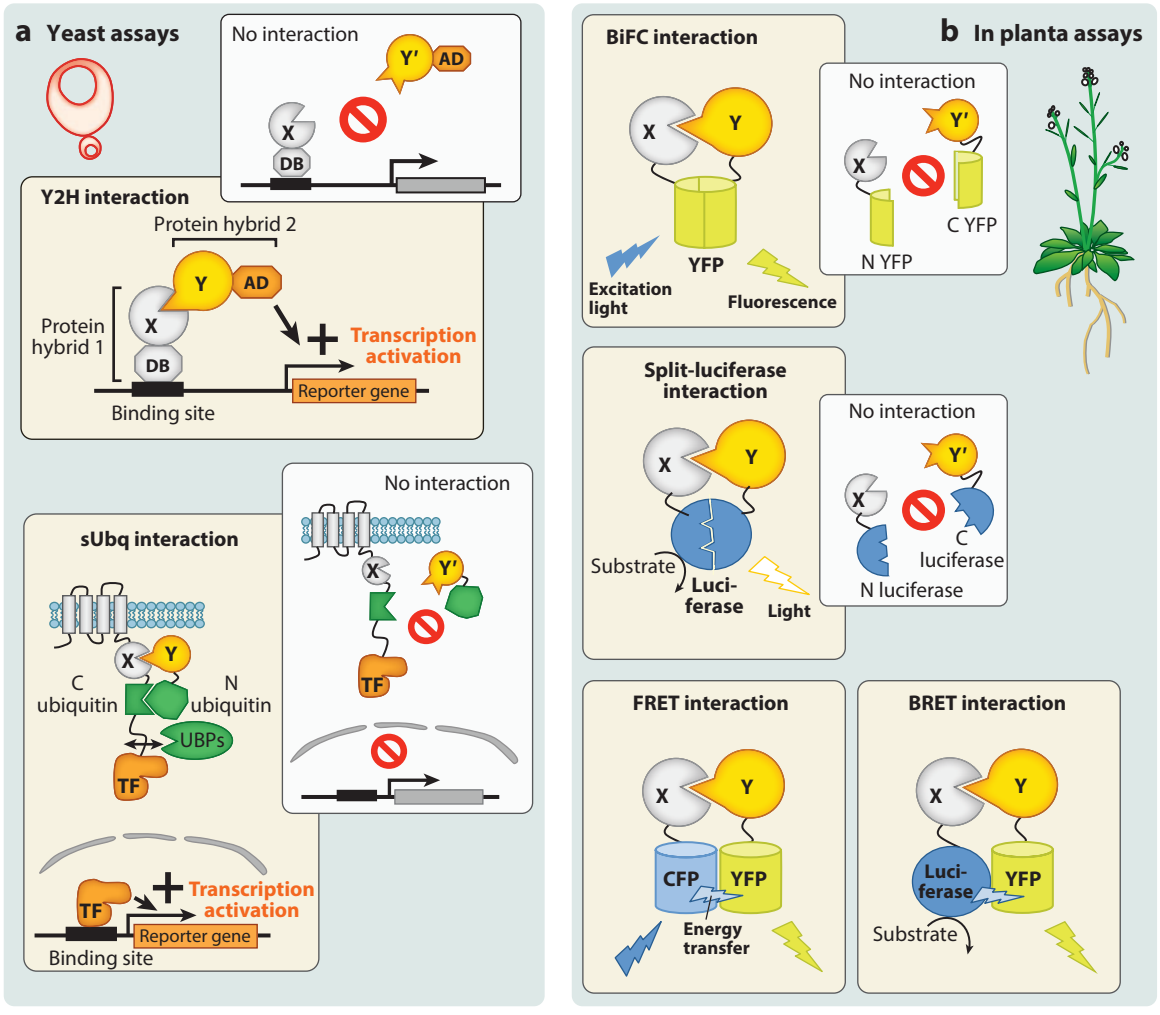


Figure 1

The *Arabidopsis* genome represented as an iceberg. Interactions between genes of largely unknown function (the hidden part, constituting approximately 90% of the genome) and those that have been functionally characterized (the tip) aid in understanding the functions of unknown proteins. Furthermore, any interaction between plant proteins and symbiotic or pathogen organism proteins will also increase understanding of the plant proteome.



assays and the role of benchmarking, see 9, 78, 90.)

Several systematic and stringently controlled experimental and bioinformatics studies have recently demonstrated that Y2H can produce data of excellent quality, often surpassing that of other assays (11, 48, 146, 158, 160). Especially important is the elimination of both constitutive and spontaneous autoactivators—usually constructs containing the DNA-binding domain that activate the reporter in the absence of an interaction (for more detail, see 29). Conversely, the false-negative rate of all assays (11) has been recognized as a major limiting factor in the construction of complete global interactome maps. This limitation can be overcome by using different orthogonal interaction mapping platforms to repeatedly screen a particular search space (9).

In plants, Y2H has been the predominant method for identifying protein interactions. Several studies on the scale of protein families or signaling modules have been conducted to address specific questions regarding the connectivity and evolutionary trajectory of the floral development network (28, 50, 71, 82, 121) and the diverse functional roles of the transcriptional corepressor TOPLESS (17), or to reconstruct network dynamics and identify plant cell cycle modules (5). These examples illustrate some of the biological questions that can be addressed by even moderately sized systematic studies.

The first systematic, experimental proteome-scale interactome map for plants, *Arabidopsis* Interactome 1 (AI-1), was published in 2011 (2). Within this high-quality network map, ~6,200 interactions connect ~2,700 proteins. In addition to doubling the number of known interactions for *Arabidopsis* and providing hypotheses about the molecular functions of several thousand unknown proteins, the systematic nature of AI-1 enabled the discovery of network communities that function in common biological processes and shed light on the subfunctionalization process of paralogs after gene duplication.

In summary, systematic data sets in which both positive and negative results are meaningful provide information that can result in more refined and better-supported hypotheses compared with only analyzing the interaction partners of a protein of interest.

Split-ubiquitin assay. The sUbq system was developed to investigate protein interactions

Positive and random reference set (PRS and RRS): controls containing reliable interaction pairs (PRS) and random protein pairs (RRS), used for assay calibration

***Arabidopsis* Interactome 1 (AI-1):** the first systematic, experimental proteome-scale interactome map for plants, containing ~2,700 proteins and ~6,200 interactions

Figure 2

Protein interaction assays. (*a*) Yeast assays. In a yeast two-hybrid (Y2H) assay (*top*), the bait (X) is genetically fused to a DNA-binding domain (DB), whereas the prey (Y) is genetically fused to a transcriptional activation domain (AD). A physical interaction between the bait and prey reconstitutes a functional transcription factor (TF), which leads to activation of a reporter gene, resulting in yeast growth or lacZ expression (*orange box*). In the absence of a physical interaction between the bait and prey, the reporter remains inactive. In a split-ubiquitin (sUbq) assay (*bottom*), the membrane-bound bait is genetically fused to a fragment of ubiquitin, which in turn is attached to a TF. The bait, or bait library, is genetically fused to the complementary ubiquitin fragment. A physical interaction reconstitutes the ubiquitin, which leads to recognition by ubiquitin-directed proteases (UBPs). The cleaved-off TF can relocate to the nucleus and activate a reporter gene. (*b*) In planta assays. In split yellow fluorescent protein (YFP) assays, also referred as to biomolecular fluorescence complementation (BiFC) (*top*) and split-luciferase (*middle*) assays, the bait and prey are genetically fused to inactive fragments of the reporter protein itself. Functional interaction between the bait and prey reconstitutes the reporter protein, enabling the detection of fluorescence and luciferase activity, respectively. In fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) (*bottom*), the physical proximity between a FRET or BRET donor (respectively) and an acceptor is mediated by physical interaction between the bait and prey. This can result in resonance energy transfer, which is detectable by appropriate instruments. (*c*) In vitro nucleic acid programmable protein array (NAPPA). In this method, plasmids encoding the bait fusion proteins are spotted on derivatized glass slides along with a suitable capture agent, e.g., antibodies or small molecules. Addition of a cell-free coupled transcription and translation protein expression system results in protein expression and subsequent bait capture by the capture agent. Addition of a plasmid for a differently tagged prey protein, or subsequent addition of this plasmid to the developed array, allows the prey to bind to the cognate bait, where it can be detected using immunological methods. Additional abbreviation: CFP, cyan fluorescent protein.

involving transmembrane proteins (103, 129) (**Figure 2*a***, bottom). Assembling screening constructs requires more individual tailoring to the proteins of interest because the N or C terminus may be on the cytoplasmic face of the membrane. As a genetic yeast-based assay, sUbq is sensitive to similar artifacts as Y2H, and controls must be implemented to remove these. In a systematic large-scale study of *Arabidopsis* transmembrane proteins, approximately one-third of baits autoactivated and needed to be removed from the screen (65); similar numbers were previously observed for yeast baits (91).

Systematic mapping of the network formed by the 490 *Arabidopsis* baits against ~4,000 preys gave a network of 343 interactions among 179 proteins (65). Subsequently, a set of 52,310 putative interactions among 414 proteins was filtered using statistical and bioinformatics approaches to yield a high-quality network of 541 interactions among 239 proteins (20). Given that fewer interactions are known for membrane proteins, quality control is generally more difficult for new data sets involving membrane proteins.

Affinity purification–mass spectrometry (AP-MS): affinity purification of protein complexes coupled to mass spectrometry for the identification of purified proteins

Immunoprecipitation: technique that immunoprecipitates protein complexes using immobilized antibodies against an epitope of the protein of interest

Tandem affinity purification (TAP): technique in which protein complexes are isolated in two consecutive purification steps, increasing the specificity

Protein complex purification from plants. The emergence of ultrasensitive mass spectrometry (MS) has fueled the development of interaction analysis methods based on biochemical purification of protein complexes. Until recently, complexes were isolated by combining sucrose density-gradient centrifugation, gel filtration, or ion-exchange chromatography. The *Arabidopsis* 26S proteasome, for example, was isolated through sequential anion exchange and size exclusion chromatography (157). However, these nongeneric methods are often characterized by a high false-positive rate. The most widely used approaches at present are based on affinity purification combined with MS (AP-MS). The classical affinity-based approaches use specific antibodies against a protein of interest (immunoprecipitation), whereas more generic approaches rely on the fusion of the target protein to an affinity handle (10).

Immunoprecipitation. During immunoprecipitation, a protein complex is purified from a cell lysate with an immobilized antibody against a known component of the complex (**Figure 3*a***). The main advantage of this method is that the endogenous protein is analyzed. Because plant antibodies are scarce and purification conditions need to be optimized for each complex, however, immunoprecipitation is not applied in a high-throughput manner. The *Arabidopsis* Mediator complex, a central coregulator of transcription, was isolated through a combination of ion chromatography and immunoprecipitation (3). In other AP-MS approaches, protein extracts are incubated not with immobilized antibodies but rather with recombinant bait proteins to trap interacting proteins. Such an approach allowed the mapping of client proteins of two 14-3-3 isoforms in *Arabidopsis* developing seeds (134).

Tandem affinity purification. One of the most powerful methods for the isolation of protein complexes is tandem affinity purification (TAP) (**Figure 3*b***). The original TAP tag (114) consists of a double-protein-A domain and a calmodulin-binding peptide separated by a tobacco etch virus (TEV) protease cleavage site. This tag enables two consecutive purification steps, increasing the specificity. The first report of TAP in plants utilized the plant-optimized improved TAP (TAPi) tag, transiently expressed in tobacco leaves, to identify interactors of the glucocorticoid receptor (115). This tag was further used in transgenic rice plants to map interactions around protein kinases (116, 117). TAP from stably transformed plants was first demonstrated with the alternative TAP (TAPa) tag, in which the calmodulin-binding peptide is replaced with a 9-Myc/6-His tag and the TEV site is replaced with a low-temperature active rhinovirus 3C protease site (120). This tag enabled characterization of the *Arabidopsis* COP9 signalosome, and a similar strategy was followed to map 14-3-3 client proteins (18).

For *Arabidopsis* cell cultures, a streamlined TAP platform is also available (141, 143),

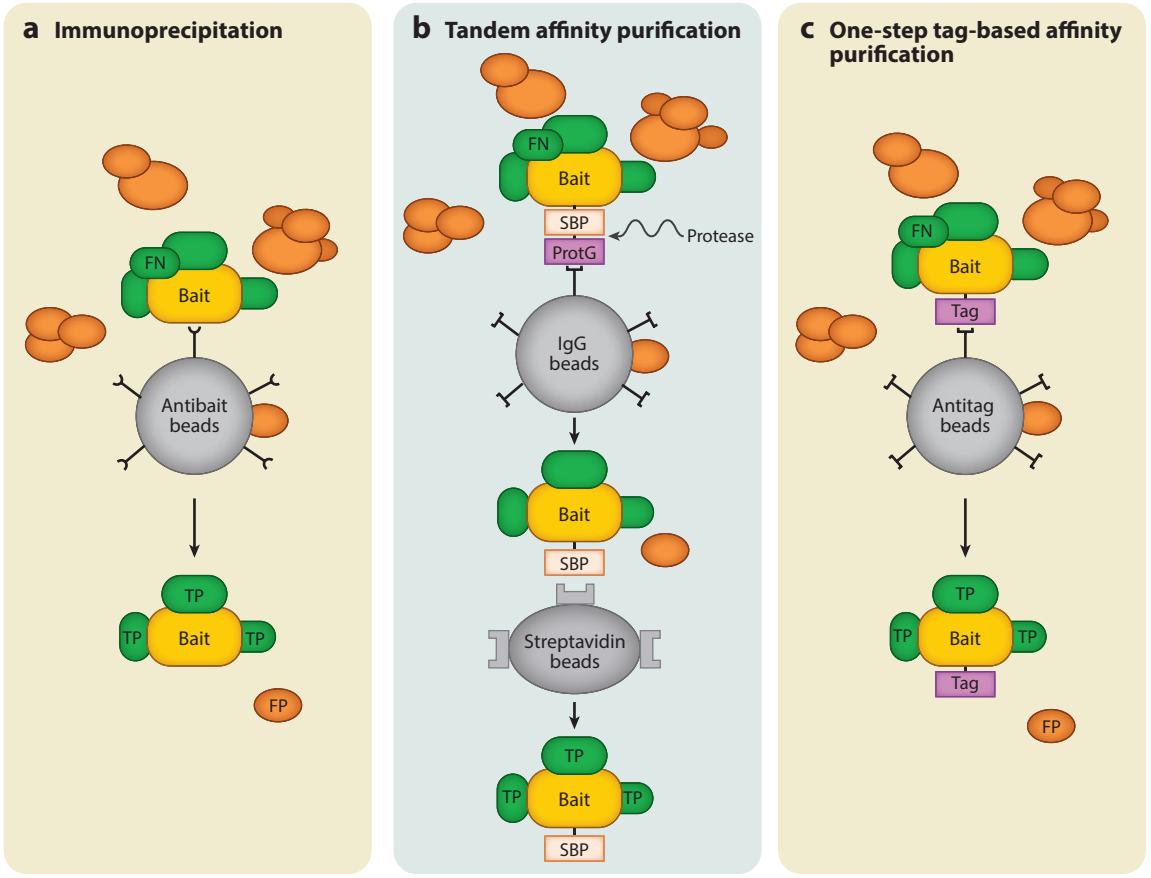


Figure 3

Strategies for complex purification from plants. (*a*) Protein complexes immunoprecipitated with immobilized antibodies directed against the bait protein of interest. (*b*) GS-tag-based tandem affinity purification of protein complexes. In the first step, complexes are isolated through binding of the protein-G (ProtG) tag to immobilized immunoglobulin G (IgG) and then specifically eluted by addition of tobacco etch virus (TEV) or rhinovirus 3C protease, depending on the protease cleavage site that is present. In the second step, complexes are isolated through binding of the streptavidin-binding peptide (SBP) tag to immobilized streptavidin and elution with (desthio)biotine. (*c*) One-step tag-based affinity purification of protein complexes. The bait protein is fused to a single affinity handle—e.g., green fluorescent protein (GFP)—allowing affinity purification in a single step. Additional abbreviations: FN, false negative; FP, false positive; TP, true positive.

allowing fast selection of transformed cells and unlimited supply of biomass. This platform was used to map a cell cycle interactome of 857 PPIs (142) and can also be used to study other basic cellular processes. TAP was further boosted through the implementation of the GS tag, which is superior to the original TAP tag in terms of both specificity and complex yield (144). The GS tag, which combines a double-protein-G domain with a streptavidin-binding peptide (**Figure 3*b***), allowed pull-down of transcription factor complexes involved in jasmonate signaling (101, 141).

In conclusion, a wide variety of powerful TAP strategies are available for unraveling protein networks in plants. Nevertheless, weak or low-abundance PPIs might be missed owing to altered binding kinetics upon extract preparation, long purification times, or low purification yields.

GFP: green fluorescent protein

Nucleic acid programmable protein array (NAPPA): a specific protein array technology that allows the detection of interactions between proteins expressed in vitro

One-step tag-based affinity purification. Many affinity handles are now available for AP-MS methods based on a single purification step (**Figure 3*c***). One major pitfall is the relatively high false-positive rate owing to the copurification of abundant or promiscuous proteins. Stable isotope labeling by amino acids in cell culture (SILAC)–like methods are not readily applied in plants. Several groups have developed statistical approaches to distinguish between weakly interacting proteins and background contaminations using additional MS information in relation to control runs (96).

The biotin peptide tag (Bio) was used to identify interactors of the TATA-box-binding protein in rice (161), and the Flag epitope tag was used to isolate the 26S proteasome from transgenic *Arabidopsis* plants (4). However, most one-step approaches in plants utilize the green fluorescent protein (GFP) tag, which also allows localization studies. The GFP tag was used to characterize MADS-domain transcription factors during *Arabidopsis* flower development in combination with label-free quantification (127) or to purify the nuclear pore complex (136). More technically challenging complexes, such as membrane-bound receptor complexes functioning in innate immunity (118) or brassinolide signaling (55), have also been analyzed using fluorescent tags.

Protein arrays. In principle, protein arrays are a powerful tool for both systematic and targeted interactome screening. To manufacture protein arrays, proteins are immobilized on chemically derivatized glass slides (85), where they are available for biochemical interrogation, e.g., using other proteins or modifying enzymes. Most protein arrays used to date have been produced by spotting recombinant proteins purified from heterologous systems (106). In an attractive alternative, the nucleic acid programmable protein array (NAPPA) developed by Ramachandran et al. (81, 109), the expression plasmids for proteins of interest are directly spotted on glass slides together with an antibody against an epitope that is genetically fused to the encoded proteins. Subsequent addition of a cell-free coupled transcription/translation system results in protein expression followed by capture of the expressed protein by the cospotted antibody (**Figure 2*c***). This platform is currently being developed for plant proteins (J. Ecker, personal communication), and a plate-adapted version has been used as a validation assay for quality control of AI-1 (2) (see below).

Protein arrays have been used to produce kinase-substrate data sets, a particular form of PPI. In plants, the substrates of mitogen-activated protein kinases (MAPKs) have been explored, uncovering an overall structure similar to that found in other biological modules (106). In this study, most kinases had few substrates, suggesting specific functions, whereas other, more versatile kinases phosphorylated substrates with many different cellular functions. A clear advantage of protein arrays is the increased control that researchers have over the experimental conditions, along with the fact that most are in vitro assays that eliminate confounding factors and variability introduced by cellular systems. However, at this stage, the expression and purification of thousands of recombinant proteins are still challenging and represent tasks that not all labs have the resources to carry out. However, commercially available arrays make this platform interesting for module-scale investigations of protein interactions.

Methods to Validate or Study Targeted Interactions

Analysis of systematic PPI networks is a powerful approach to gain insight into the systems organization of proteomes and address biological and evolutionary questions. However, independent of how carefully a mapping experiment is performed, at least a representative subset of the data should be validated using an independent assay (29). This is a challenge because no two assays detect an identical subset of "true" interactions (9, 11, 54, 158). This challenge can be overcome by calibrating the performance of the validation assay using a PRS and RRS (11, 146, 158).

The PRS should consist of well-documented interactions from the literature (25); the RRS is composed of random protein pairs, which a priori have a low probability of being true interactors (9). When all PRS/RRS pairs are tested in the validation assay, the PRS detection rate yields the sensitivity, whereas the RRS rate reflects the assay background (9, 11). Within this window, the validation data of a random subset of the data set can be interpreted: A rate close to the PRS indicates high data quality, and a rate close to the RRS indicates low data quality (figure 1 in Reference 9; 2, 146, 158).

When focusing on particular interactions in the context of a biological study, most reviewers will require validation in planta. Again, this task can be difficult or even impossible if a single validation method is preferred. Therefore, it is often necessary to use a combination of methods and tailor the conditions of any given method to the specific proteins. Many PPI methods have been developed during the past 10–15 years, and most of them are common to both animals and plants.

In planta split methods. The split methods, also called protein-fragment complementation assay (PCA) methods, encompass different approaches based on a simple scheme. The bait and prey proteins to be tested for interaction are fused to two fragments of a reporter protein, neither of which by itself has reporter activity. During the interaction, the bait and prey proteins bring together the two fragments and reassemble an active protein (58) (**Figure 2*b***, top). Although PCA methods can be performed in vitro, they are more commonly used in vivo. Ten families of reporters have been developed in mammals (reviewed in 58). Among these reporters, with the exception of a single study that used dihydrofolate reductase in plants (132), GFP and its variants have been commonly used, and split luciferase has been shown to be efficient for studies of PPI dynamics.

A PCA method that uses GFP or variants as reporters is referred to as bimolecular fluorescence complementation (BiFC) (58). First described in bacteria (40), BiFC was then used in mammals (46) and subsequently adapted to plants (7, 138, 150). Even though the GFP fluorescence signal can be observed using a regular epifluorescence microscope, a confocal microscope is often needed in plant cells because of the background autofluorescence of photosynthetic pigments (7). The main caveat associated with BiFC is that the interactions stabilize by refolding, which can cause artifacts (46, 86). However, this can sometimes be turned into an advantage because it facilitates the study of transient interactions, as has been shown for kinase substrates in plants (108).

BiFC is generally combined with transient expression in *Nicotiana benthamiana* or *Arabidopsis*, but it has been used in many other plant systems and/or with stable expression (23). Several binary plasmids for plant transformation have been developed for regular or Gateway-based cloning (23, 38, 89, 150). In addition, some interesting developments of the BiFC methods have been described in animals, including multicolor BiFC (47), ternary BiFC (124), and interaction screening using BiFC coupled to fluorescence-activated cell sorting (FACS) (92). Among these, only multicolor BiFC has been adapted to plants (69, 148).

The use of luciferase as a reporter for PCA methods was first described in mammal cells in 2004 (84) and was subsequently adapted to plants using either *Renilla* (35) or firefly (19) luciferase. Luciferase activity is detected using a luminometer in the dark, and the throughput can be scaled up using a microplate luminometer (72). This method is particularly useful for plants because autofluorescence cannot be excited in the dark. Split luciferase was shown to be the most sensitive and dynamic PPI detection method, and is able to detect protein dissociation in both animals (83) and plants (72).

Coimmunoprecipitation. Coimmunoprecipitation is a widely used method to validate PPIs. In this technique, the bait protein is precipitated with an immobilized antibody, and the interacting partners are analyzed by western blotting or MS. Several strategies can be followed, using antibodies against

Protein-fragment complementation assay (PCA): generic term for an assay based on reconstituting a functional reporter from two inactive fragments

Bimolecular fluorescence complementation (BiFC): protein interaction assay based on the reconstitution of a fluorescent YFP from two nonfluorescent fragments

FRET: fluorescence resonance energy transfer

BRET: bioluminescence resonance energy transfer

Well NAPPA (wNAPPA): NAPPA-derived binary assay performed in microtiter plates

the endogenous proteins or against tagged fusion proteins either transiently or stably expressed. Coimmunoprecipitation has been used to demonstrate ligand-induced interactions involved in innate immunity (118). Coimmunoprecipitation experiments usually generate significant background, however, and it is therefore important to perform appropriate negative controls (110).

Other assays for validation and targeted interaction analysis, such as luminescence-based mammalian interactome mapping (LUMIER) and mammalian protein-protein interaction trap (MAPPIT), have been used mostly in the context of animal and yeast interactome mapping (discussed in 9).

Resonance energy transfer methods. When the physical distance between two fluorescent proteins is small (<100 Å), energy can be passed from one to the other by a phenomenon called resonance energy transfer. This optical property can be used to demonstrate close physical distance between either two fluorescent fusion proteins [in fluorescence resonance energy transfer (FRET)] (36) or one bioluminescent donor fusion protein and one fluorescent acceptor protein [in bioluminescence resonance energy transfer (BRET)] (155) (**Figure 2*b***, bottom). Both methods have been applied to the study of PPIs in animal and plant cells (36, 155).

FRET results in a decrease in the donor fluorescence intensity and an increase in the acceptor fluorescence intensity. The actual FRET measurements are rather complex because the decrease and increase in fluorescence depend on several local parameters (36). In contrast, the fluorescence lifetime of the donor chromophore decreases with FRET independently of chromophore concentration and other local parameters, and is used to measure FRET efficiency in the fluorescence lifetime imaging microscopy (FLIM) method (36).

Well NAPPA. Well NAPPA (wNAPPA) is a modification of NAPPA that is performed as a pairwise assay in microtiter plates, a setup that is more suitable for validating data sets. The assay is similar to the NAPPA array except that the protein expression is done in microtiter wells and the plasmids are not spotted but rather stay in solution. wNAPPA was used for quality control of the AI-1 data set by testing a set of 260 new interactions of unknown quality and comparing their retest rate with the performance of the assay against an *Arabidopsis* PRS/RRS set (for details, see below and References 2, 9, and 29). In this benchmarking experiment, wNAPPA had an overall sensitivity of ~20%, with some background (2). The retest rate of AI-1 was statistically indistinguishable from the PRS detection rate, thus demonstrating its high data quality.

Comparison of the Methods

Comparing methods is challenging and has led to passionate debates for a long time. Whereas cost and throughput are quite easy to quantify, comparisons of sensitivity and accuracy are more challenging. The challenges are aggravated by the fact that performance parameters are not inherent properties of any technology, but rather reflect the stringency and care with which a given assay is implemented, executed, and scored. Moreover, in data set comparisons, additional confusion can arise because different proteins are tested and incomplete assay sensitivity is often ignored. Recently, however, the development of standardized interaction reference sets for different organisms and their use in characterizing several interaction assays by independent labs have begun to overcome some of these challenges, and have yielded some surprising results (11, 21). For example, in a comparison of five binary interaction assays using human PRS/RRS sets, all assays had similar overall sensitivities (between 20% and 40%), but the profile of detected positive control PRS interactions was unique for each assay (11).

Throughput. Yeast methods (Y2H and sUbq) are currently the only available true high-throughput assays that enable interactome mapping at the full-proteome level (**Figure 4*a***,

Table 1). This process has been started for *Saccharomyces cerevisiae* (32, 51, 52, 139, 158), *Caenorhabditis elegans* (6, 74, 125), *Drosophila melanogaster* (41), humans (119, 131), and recently *Arabidopsis* (2). In addition, the sUbq system has the potential to be set up to the proteome scale (65), with the possible advantage of having a detection profile that may be complementary to Y2H. However, the complementarity remains to be experimentally verified, e.g., using common reference sets. In planta methods such as TAP and split luciferase are medium-throughput methods, mainly because of the bottleneck of plant transformation. In theory, protein arrays based on in vitro transcription/translation systems could be optimized to reach a high throughput.

Completeness of the search space. In terms of the testable search space, the cloning step is a key bottleneck of most PPI methods. Indeed, in yeast, split, resonance energy transfer, and protein array methods, the search space is restricted to the square of the number of available ORFs (146). For example, whereas Y2H achieves high throughput, because ORFs are available for only half of the *Arabidopsis* proteome, the *Arabidopsis* interactome mapping project is currently restricted to one-fourth of the whole interactome matrix (2). In contrast, affinity purification methods need only one of the partner proteins to be cloned, and in theory these methods allow the identification of interacting proteins in the whole interactome. Practically, however, the search space is further restricted to the proteins expressed in the chosen tissues. Moreover, genome annotation further limits the completeness—i.e., if the coding sequence is not correctly annotated in the genome, the corresponding peptides will not be identified.

Accuracy (false positives). Although accuracy is one of the most important qualities of a PPI method, it is difficult to quantify and to compare between methods. It is helpful to distinguish between technical and biological false positives. Technical false positives result from assay artifacts and correspond to protein pairs that do not actually interact. When sources of artifacts are known, they can be removed by appropriate controls. In this respect, methods for which more experience about artifacts has accumulated can be expected to produce higher-quality data than new methods for which a smaller amount of data is available. In a comparison of several interaction assays using identical PRS/RRS sets containing ~100 protein pairs each, Y2H was the only assay that had no background; other methods, such as split yellow fluorescent protein (YFP), LUMIER, and wNAPPA, detected 1–4 random protein pairs (11). Although 100 reference protein pairs is too small a number to obtain proteome-scale significance, this approach enables standardized comparison of the accuracy and sensitivity of different assays.

YFP: yellow fluorescent protein

In addition, the accuracy of each method depends on the stringency of the protocol. All technologies can produce high false-positive rates if the protocol is not implemented stringently enough and good negative controls are not used (78). Empty vector controls are commonly not suitable as negative controls as they do not appropriately reflect the biochemical complexity of fusion proteins in the assay. PRS/RRS sets are useful for optimizing assays to obtain the highest sensitivity while maintaining an acceptably low background. The acceptable background may depend on the particular application.

Biological false positives correspond with protein pairs that biophysically interact in an in vitro or heterologous assay but never have the opportunity to do so during the plant's life because they are never expressed in the same tissue, in the same subcellular localization, and/or at the same time. Obviously, exogenous expression of two proteins can find interactions that will never exist in the plant, whereas methods that require overexpression of a single fusion protein can lead to artificial interactions that do not occur under physiological conditions.

Sensitivity (false negatives). Sensitivity is one of the most important drawbacks of PPI

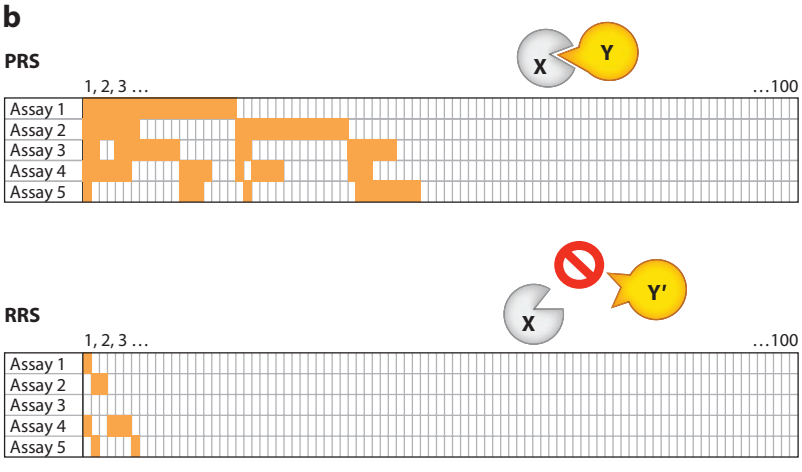
methods, and it is too often overlooked. If the stringency of a protocol is set up to limit false positives, a high rate of false negatives is usually observed. For example, a sensitivity rate of 20–40% has been observed in the benchmarking experiments of five interaction assays—i.e., 60–80% of real interactions were missed (11, 98). PRS/RRS pairs were used to show that the

a

	In planta	Heterologous / in vitro	Transient expression	Throughput	Binary interaction only	Subcellular localization of the interacting proteins	ORF cloning required	Space of search = (ORFeome size)^2	Space of search = whole proteome	Easy to set up in a regular plant molecular biology lab	Special equipment required
Y2H		X		High	X		X	X		X	Robots (for high throughput)
sUbq		X		High	X		X	X[a]		X	Robots (for high throughput)
Protein array		X		Low	X		X	X			
TAP	X			Medium			X		X[b]		Mass spectrometer
BiFC	X		X	Low		X	X	X		X	Confocal microscope
Split luciferase	X		X	Medium			X	X		X	Luminometer
Coimmuno-precipitation	X		X	Low					X[c]	X	
FRET-BRET	X		X	Very low		X	X	X			Confocal microscope

[a] Membrane protein available as ORFs × ORFs
[b] If expressed in the sample
[c] Low throughput

b



assay sensitivity of the Y2H version used to assemble AI-1 was approximately 36% (2), a value comparable to that of large-scale yeast and human screens (146, 158). These rates of false negatives make it nearly impossible to prove that two proteins do not interact.

There are numerous reasons for false negatives. The site of tagging in fusion proteins may interfere with interactions or with reporter assembly, thus decreasing the assay efficiency (7, 35, 72); this is also the underlying reason that experimental bait-prey switching often yields different results even for reliable interactions (21). Tagging may also affect the subcellular localization, or a given interaction may not be able to occur in the subcellular compartment where the assay is performed. A protein complex can be stable, whereas the isolated binary interactions are not. During purifications, weak interactions may be washed off, and overexpression of the bait may be toxic for cells. Thus, biophysical, biochemical, and steric properties of the fusion proteins; absence of important cofactors; and incorrect processing or folding of the proteins could prevent proteins from interacting (5).

Therefore, no method is inherently preferable; researchers must choose the appropriate protocol based on the objective of the experiment and the known limitations of each assay. In addition, using a combination of methods is often helpful to identify interactions more comprehensively or increase confidence in an identification (63) (**Figure 4*b***). This is similarly true, although more expensive to implement, for proteome-scale interactome mapping (9). Indeed, confirming an interaction by a second method is the best way to increase the confidence of the results (127). However, it is important to note that systematic application of this validation step also reduces mapping throughput and sensitivity. In contrast, combining all interactions obtained by two methods to add interactions increases the coverage but severely reduces accuracy (5, 16, 93, 122, 123, 140).

CONSTRUCTING AND DRAWING THE INTERACTOME

Prediction

Despite the progress of high-throughput technologies and their application in the plant kingdom, the results obtained to date are far from providing access to the hundreds of thousands of PPIs that occur in a plant cell (2). In an attempt to fill in this gap, various computational methods have been developed to predict new PPIs.

The first type of predictive method is based on the physical-chemical properties of the sequences and the 3D structures of the proteins (104). The Protein Data Bank, a unique reference for the 3D structures of proteins and

Figure 4

Comparison of protein-protein interaction methods. (*a*) The main characteristics of each method. Red indicates a drawback; blue indicates an advantage. (*b*) Conceptual representation of interaction assay benchmarking using a positive and random reference set (PRS and RRS, respectively). The top panel represents 100 PRS protein pairs corresponding to well-documented interactions from the literature (positive controls); the bottom panel represents 100 RRS protein pairs, which have an a priori low probability of interacting (negative controls). Testing all protein pairs in the five assays reveals that each assay has a unique PRS detection profile. However, under conditions of low RRS detection, all assays have a similar overall sensitivity of 20–40%. The different detection profiles and incomplete assay sensitivity necessitate calibrating validation assays for sensitivity and background. Without such calibration, the limited assay sensitivity of the validation assay would result in the incorrect conclusion that a smaller fraction of the new data set can be reproduced and hence lead to an overestimation of the false-positive rate. Additional abbreviations: BiFC, bimolecular fluorescence complementation; BRET, bioluminescence resonance energy transfer; FRET, fluorescence resonance energy transfer; ORF, open reading frame; sUbq, split ubiquitin; TAP, tandem affinity purification; X, bait; Y, prey; Y2H, yeast two-hybrid. Data are from References 11 and 21.

Table 1 Main protein-protein interaction (PPI) resources for plants: data sets containing more than 50 interactions

Data set/reference	Plant	Object of investigation	Space of search	Number of proteins in the network	Number of interactions in the network	PPI method
MIND 0.5	*Arabidopsis*	Membrane	1,043 proteins × 3,307 proteins	2,300	24,777	sUbq
AI-1 (2)	*Arabidopsis*	Systematic	~8,000 proteins × ~8,000 proteins	2,661	5,664	Y2H
Mukhtar et al. (94)	*Arabidopsis*	Pathogens	552 effectors × ~8,000 *Arabidopsis* proteins	926	1,358	Y2H
Popescu et al. (106)	*Arabidopsis*	MAPK	10 MAPKs × 2,158 proteins	580	1,280	Protein arrays
Klopffleisch et al. (63)	*Arabidopsis*	Protein-G complex	7 protein-G-related proteins × 9 cDNA libraries	434	1,058	Y2H
Van Leene et al. (142)	*Arabidopsis*	Cell cycle	102 cell cycle proteins × whole of proteome	393	857	TAP-MS
Brizard et al. (12)	Rice	Rice virus	Proteome	224	No data	Protein purification
Popescu et al. (105)	*Arabidopsis*	Calmodulins	7 calmodulins × 1,133 proteins	180	716	Protein arrays
Causier et al. (17)	*Arabidopsis*	TOPLESS	5 TOPLESS proteins × library	No data	655	Y2H
Schoonheim et al. (122)	Barley	14-3-3 family	5 14-3-3 isoforms × cDNA library	155	~500	Y2H + TAP-MS
Boruc et al. (5)	*Arabidopsis*	Cell cycle	58 cell cycle proteins × 58 cell cycle proteins	58	357	Y2H + BiFC
Lalonde et al. (65)	*Arabidopsis*	Membrane	445 proteins × 206 proteins	179	343	sUbq
de Folter et al. (28)	*Arabidopsis*	MADS box	~100 MADS-box proteins × ~100 MADS-box proteins	77	272	Y2H
Chang et al. (18)	*Arabidopsis*	14-3-3 family	1 14-3-3 isoform × whole of proteome	131	130	TAP-MS
Swatek et al. (134)	*Arabidopsis*	14-3-3 family	2 14-3-3 isoforms × whole of proteome	106	129	AP-MS
Leseberg et al. (71)	Tomato	MADS box	22 LpMADS-box proteins × 23 LpMADS-box proteins	21	119	Y2H
Zhong et al. (161)	Rice	TATA-box protein complex	1 TATA-box protein × whole of proteome	86	85	AP-MS
Richardson et al. (113)	*Arabidopsis*	Endosome sorting complex	28 ESCRT proteins × 28 ESCRT proteins	18	77	Y2H
Seo et al. (123)	Rice	Stress response	Combination of subinteractomes	84	77	Y2H + BiFC
Immink et al. (50)	Petunia	MADS box	23 PhMADS-box proteins × 23 PhMADS-box proteins	23	64	Y2H
Hunt et al. (49)	*Arabidopsis*	Polyadenylation	28 candidate proteins × 28 candidate proteins	28	56	Y2H
Ehlert et al. (30)	*Arabidopsis*	bZIP	17 bZIP proteins × 11 bZIP proteins	9	51	Y2H
Ciannamea et al. (22)	Ryegrass	MADS box	4 LpMADS box proteins × 24 LpMADS box proteins	27	51	Y2H

Abbreviations: AP, affinity purification; BiFC, bimolecular fluorescence complementation; MS, mass spectrometry; sUbq, split ubiquitin; TAP, tandem affinity purification; Y2H, yeast two-hybrid.

protein complexes (145), is the starting point for methods that explore the experimentally characterized protein interfaces to highlight their features and then look for them in proteomes in order to predict domain-domain interactions (31, 60, 62). Protein-docking algorithms also constitute an important class of methods to predict interaction sites and protein-protein binding interfaces (128); these algorithms are regularly evaluated in the framework of the Critical Assessment of Prediction of Interactions (CAPRI) initiative (70). Their performance remains limited, especially regarding transient (unstable) protein complexes and given the relatively low number of accurate 3D models available. This latter point is an actual bottleneck for the structure-based predictive approaches, which is why the majority of the tools to predict PPIs at a large scale are independent of structural data and instead are predominantly based on comparative genomics and homology.

The second type of predictive method builds on the fact that protein complexes are structurally and functionally conserved during evolution. Consequently, the joint presence/absence of homologs of a pair of proteins in a great number of species can be an indication of functional relationships. This method of phylogenetic profiling (102), which does not require experimental results, is applied mainly in the prokaryote kingdom, for which a huge number of fully sequenced genomes are available (126).

The conservation of interactions during speciation events is also exploited to transfer experimentally characterized PPIs between organisms. This inference of interactions is based on the definition of orthologs for each protein partner, resulting in the identification of pairs of "interologs," as defined by Walhout and collaborators (149). In this context, the interactome prediction becomes a problem of predicting orthology relationships between model species (to which high-throughput experimental approaches have been applied) and newly sequenced and annotated genomes. Sequence comparisons and sometimes gene context (synteny analysis) are used to define groups of orthologous genes/proteins (64). For instance, the 76,585 rice PPIs of the Predicted Rice Interactome Network (PRIN) resource (162) have been deduced from interologs between yeast, nematode, fruit fly, human, *Escherichia coli*, and *Arabidopsis* based on ortholog clusters predicted by the InParanoid algorithm (100). This algorithm has also been used by Geisler-Lee and collaborators (39) to define interologs between yeast, nematode, fruit fly, and human, leading to the prediction of an *Arabidopsis* interactome of 19,979 PPIs, which have been scored according to the number of supporting species and methods. In combination with the structural domain-based method of *i*Pfam (34), the interolog strategy has been applied to predict PPIs between plants and pathogens, e.g., between *Arabidopsis* and *Ralstonia solanacearum* (75) and between rice and *Xanthomonas oryzae* (61).

With the aim to enrich and better functionally characterize the *Arabidopsis* interactome predicted through the definition of orthologs, De Bodt and collaborators (27) have applied filters and computed confidence values in considering the colocalization of the candidate protein partners and the coexpression of the corresponding genes (Pearson correlation). In this way, they have obtained a "filtered" interactome of 18,674 PPIs. The resulting network has been analyzed to define highly connected protein clusters, assuming that they may gather protein partners involved in the same biological function or protein complex. The functional annotation of the clusters (Gene Ontology enrichment) shows the relevance of this approach (27).

The third type of predictive method, referred to as combiners or metapredictors, is based on the integration of functional data. These methods combine a variety of predicted and experimental evidence (i.e., heterogeneous data) to propose interactions. Most of them use machine learning methods such as neural networks, Bayesian classification, linear regression, or support vector machines (SVMs). All are based on a training step to extract knowledge from a clean validated data set (the gold

standard positive set) and calculate confidence values correlated to the amount of evidence that supports each predicted interaction.

Such integrative approaches were first applied to yeast and human interactomes (53, 112). In plants, the amount of available functional data focuses the efforts on *Arabidopsis*. The *Arabidopsis thaliana* Protein Interactome Database (AtPID) predicted interactome (containing 23,396 pairs) has been generated by a Naive Bayesian approach integrating interologs, gene expression, phylogenetic profiles, genomic context, gene split/fusion [the Rosetta stone method developed by Marcotte & Marcotte (87)], protein domains, Gene Ontology annotation, and text-mined interactions (154). Furthermore, in a recent update, the AtPID network was enriched by the inclusion of mutant phenotypes, protein phosphorylation, proteomics, and transcriptional regulation data (73). Arguing that the SVM approach has been evaluated as a better supervised learning method and that it tolerates the integration of correlated features (in contrast to Naive Bayesian models, which treat data sets as independent), Lin and collaborators (79, 80) have used it to merge the same types of data (14 features regarding annotation, expression, and interacting domains) and predict 145,494 PPIs in *Arabidopsis* [constituting the Predicted *Arabidopsis* Interactome Resource (PAIR) data set].

The evaluation of all these predicted interactomes is a central but difficult problem. Authors exemplify their reliability by selecting a few situations where the interactions are highly coherent, with biological knowledge from the literature. At a larger scale, experimentally characterized interactions (not included in the training set) are sometimes used for comparisons with the predictions. According to the resources, the deduced sensitivity ranges from 9% to 20%, which means a minimum false-negative rate of 80% (68, 79). PAIR, the largest predicted interactome in *Arabidopsis*, shares only 368 interactions with the 6,207 experimental PPIs of the Y2H AI-1 interactome (2). Furthermore, the poor overlaps between different predicted interactomes illustrate that the selection of the experimental resources used to build the gold standard positive data set and the selection of the computational approaches (linked to the scoring method) have a deep impact on the results (27, 104). Regarding the specificity, the absence of a gold standard negative data set prevents the estimation of false-positive rates. To bypass this limitation, the topology (connectivity analysis) of the predicted interactome network and the intersection between the predicted pairs and independent experimental data (expression, subcellular localization) are sometimes used as indirect quality indicators (27, 39).

Database and Visualization

Mindful that the organization, integration, curation, and homogeneity of interactions are the limiting points for their use by the scientific community, the first interaction providers have created the International Molecular Exchange (IMEx) consortium. This international effort has defined the minimal information for reporting molecular interaction experiment (MIMIx) guideline and proposed a web portal dedicated to the curation of protein interactions (99). In this initiative, the plant PPIs, represented by the model *Arabidopsis*, reach 9.7% of the IMEx data set. The main repositories for plant PPIs are IntAct (59) and the Biological General Repository for Interaction Datasets (BioGRID) (130). Furthermore, the *Arabidopsis* Information Resource (TAIR) plays a major role in centralizing *Arabidopsis* protein interactions (66). Most of the experimental and predicted interactome resources have their dedicated information system, but metadatabases have also been developed to give access to several resources through a single-query interface. **Table 2** provides a (nonexhaustive) list of databases that host plant protein interactions.

Interactome data are represented as an undirected graph where each protein is a node and physical interactions are symbolized by edges. The degree of a node defines the number of interactions that the corresponding protein has. Typical biological networks are characterized by a sparse topology with a limited number of

Table 2 Main protein-protein interaction (PPI) resources for plants: main databases that host plant protein interactions

Database	URL	Plant(s)	PPI type	Reference
IntAct	**http://www.ebi.ac.uk/intact**	All	Experimental	Kerrien et al. (59)
BioGRID	**http://www.thebiogrid.org**	Mainly *Arabidopsis*	Experimental	Stark et al. (130)
DIP	**http://dip.doe-mbi.ucla.edu/dip**	*Arabidopsis*	Experimental	Xenarios et al. (153)
MINT	**http://mint.bio.uniroma2.it/mint**	*Arabidopsis*	Experimental	Licata et al. (77)
APID	**http://bioinfow.dep.usal.es/apid**	*Arabidopsis*	Experimental	Prieto et al. (107)
PAIR	**http://www.cls.zju.edu.cn/pair**	*Arabidopsis*	Predicted	Lin et al. (79)
PRIN	**http://bis.zju.edu.cn/prin**	*Oryza*	Predicted	Gu et al. (44)
AtPID	**http://www.megabionet.org/atpid**	*Arabidopsis*	Predicted	Li et al. (73)
*i*Pfam	**http://ipfam.sanger.ac.uk**	All	Predicted	Finn et al. (34)
BAR	**http://bar.utoronto.ca**	*Arabidopsis*, *Oryza*	Experimental and predicted	Geisler-Lee et al. (39)
STRING	**http://www.string-db.org**	All	Experimental and predicted	Szklarczyk et al. (135)
AtPIN	**http://bioinfo.esalq.usp.br/atpin**	*Arabidopsis*	Experimental and predicted	Brandão et al. (8)

hubs, i.e., proteins having a high degree of connectivity. This so-called scale-free topology is observed in both experimental and predicted PPI networks (44). Many bioinformatics tools have been developed to display and explore this network representation (133), but only a few are widely used. Among these tools, Cytoscape has proven to be the most efficient and flexible solution (24).

Evolution of Interactome Networks

An important and fascinating aspect of the growing field of interactome analysis is the comparison of networks from different organisms or accessions and reconstitution of the evolutionary trajectories that led to adaptation and the appearance of new species or traits. This is illustrated by a comparison of the interactomes of yeast and *Helicobacter pylori*, in which the authors found that several specialized pathways in yeast arose through gene duplication followed by specialization of existing pathways (57). In general, however, detailed comparisons are difficult, because of the incomplete nature of all existing maps owing to the discussed limitations in sensitivity, different mapping platforms, and other factors (146).

Some high-level comparisons are possible, though. All networks appear to have a scale-free topology. Analysis of AI-1 has revealed that proteins specific to the plant kingdom are not clustered in isolated areas but rather are indistinguishable in their network connectivity from those proteins common to plants, animals, and fungi (2). AI-1 also provided the opportunity to investigate the evolutionary kinetics of network rewiring after gene duplication, showing that most interactions common to a duplicated protein pair are rapidly lost after duplication. However, in contrast to exponential loss kinetics, a significant fraction of common interactions are maintained over long evolutionary timescales. Together, these observations support the notion that selective pressure acts on the network; temporary release of this pressure after gene duplication allows rapid network rewiring, but as both duplicated proteins assume new or specialized functions, the pressure retightens.

A comparison of interactome network densities was possible for yeast, human, and plants based on interactome sizes estimated with a quantitative framework (2, 146, 158). For plants, the size of the complete interactome network was estimated to be ~300,000 binary protein interactions among the 27,000 *Arabidopsis*

proteins. Fascinatingly, all three networks have a similar overall density of ~1–3 interactions per 1,000 pairwise protein combinations. Although this finding must be considered preliminary, it indicates that biological networks may share some universal properties.

USING THE PLANT INTERACTOME

Inference of Function and Integration with Other -Omics Data

Interactome data are a major component of biological system approaches, which are based on the integration of numerous quantitative and qualitative biological resources to drive new hypothesis formulation (76). The obtained networks highlight new relationships between genes or proteins, and highly connected substructures might point to putative functional groups, i.e., operational units of physiological functions. This assumption allows significant progress in the functional annotation of genes. Indeed, pairwise connection can be used to infer function between known and unknown genes, thus going beyond the limited inference by sequence similarities. This guilt-by-association strategy has been applied and positively evaluated in different biological contexts, including metabolic pathways (79, 80, 154), drought response (67, 68), and floral transition (45).

With the aim of improving the reliability and gene coverage of the inference of knowledge by association, integrative approaches merging computational results and experimental -omics data have been developed. For instance, the *Arabidopsis* Reactome (137) and VirtualPlant (56) resources work on the merging of molecular pathways, transcriptomics, PPIs, and literature-mined functional data. Important conceptual efforts have aimed at representing such heterogeneous data to aid the prediction of master regulators and the understanding of pathway interactions in various biological processes. Furthermore, the definition and integration of orthologous relationships between sequenced plant genomes facilitate translational research by knowledge transfer between species. The power of such an integrative way to propose relational annotations is illustrated by the AraNet network, which consists of more than 1 million functional links encompassing 73% of the *Arabidopsis* genes (67, 68). CORNET (for correlation networks) is also an ambitious initiative, with a high level of integration in *Arabidopsis* (14 PPI data sets and more than 20 different resources) and some initial applications in maize (26).

Lastly, the major reference in data integration (in terms of methods, resources, and graphical display) is probably the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (135). This database merges and interprets several data points to predict functional linkages and physical interactions between proteins. It takes extensive advantage of genome context methods (gene neighboring, gene split/fusion, and phylogenetic profile), clusters of orthologous genes, coexpression data, and experimental and predicted knowledge extracted from biological databases and the literature to assign a probabilistic confidence score to each functional association. Initially focused on prokaryotes, STRING now manages more than 1,000 sequenced species, including *Arabidopsis*, *Brachypodium*, *Physcomitrella*, rice, poplar, sorghum, and grapevine.

Interactions Between Plant Interactome and Pathogen Proteins

In times of climate change, pathogen migration, and intensified agriculture across the world, understanding pathogen-host interactions on the molecular level has immense practical implications beyond academic interest. Microbial pathogens subvert the plant immune system by injecting so-called virulence effectors into the plant cytosol (37, 97). Over the past decade, great progress has been made on elucidating the mechanisms by which selected model effectors modify their targets and thereby interfere with defense signaling and defense mechanisms.

The recent construction of a systematic large-scale interactome network map for

Arabidopsis allowed the investigation of effector interactions with the host proteome from a systems perspective (94). This study investigated interactions of the host proteins with effectors from the evolutionarily distant oomycete *Hyaloperonospora arabidopsidis* and the bacterium *Pseudomonas syringae*, both of which are important models for studying host-pathogen interactions. The study arrived at several important and fascinating conclusions. First, the effector interactions from both pathogens converge on only a few proteins in the host network. Together, they "target" fewer proteins than expected, and effectors from both pathogens bind to the same proteins more frequently (common targets) than expected, indicating that these are particularly important for infection. Although no enrichment of classic immune proteins could be found among either the common effector targets or all effector targets, more than 90% of the common targets could be verified genetically and exhibited an enhanced disease susceptibility or enhanced disease resistance phenotype. This extraordinary genetic validation rate illustrates the power of network analyses for uncovering facets of host-pathogen interactions that are not accessible by small-scale approaches.

Interestingly, nearly half of the *Arabidopsis* common targets are highly connected hubs in the host interactome network. Theoretical network analyses have identified such hubs as the Achilles' heel of scale-free networks (1). It was also observed that many proteins are repeatedly targeted by effectors from the same pathogen; however, the biological reason for this is currently unclear. It is reasonable that the multiple targeting by effectors of the same pathogen grew by divergent evolution out of a single effector binding to this protein (original targeting). Subsequent multiple targeting could in principle arise by effector duplication followed by too little evolutionary divergence of the new effector copies. However, the observation that the same protein is bound by several effectors from distinct and evolutionarily distant pathogens makes this explanation less likely. Another, more intriguing possibility is that the respective proteins mediate different functions in different stages of the immune response, and several effectors are needed to suppress these functions. Either way, more detailed functional studies of the respective proteins and their role in immunity are needed.

CONCLUSIONS AND FUTURE ISSUES

In the past 10 years, tremendous progress has been made in interactome network analysis. The study of plant networks has undoubtedly contributed to and profited from this development. Exciting questions can now be addressed, and at the same time the remaining technical hurdles need to be tackled.

Although the technical challenges associated with data quality have been at least conceptually solved, the low coverage of existing maps remains a problem. To address this, it will be necessary to develop new assays or improve existing ones to achieve a greater overall sensitivity. At the same time, the throughput of interaction methods needs to be improved, both to map the *Arabidopsis* network more completely and (perhaps more importantly) to map additional networks of close relatives and important crops. Taking advantage of second- and third-generation sequencing suggests a promising path toward this goal (159). The expanding data sets can then be expected to lead to a better theoretical understanding of the biophysical and biological properties that enable proteins to interact. This better understanding can be translated into improved algorithms for interaction prediction such that interactome models of reasonable accuracy and practical utility can be inferred for those crops that are less experimentally accessible.

Biologically, it will be exciting to see to what extent genetic variation quantitatively and qualitatively affects network wiring and how this in turn affects phenotypes. Likewise, the first analysis of host-pathogen interactions has provided intriguing insights into how pathogens manipulate the host networks, and similar studies in crops and with different pathogens

will undoubtedly have immense practical implications. Of particular interest is the expansion of network analysis to crops, which brings the potential to exploit this knowledge to modify existing crops in order to produce higher yields or, perhaps more importantly, produce stable yields in the context of increasing biotic and abiotic stresses. As this article was being written, the United States was being hit by the worst drought on record, and drought and wildfires were damaging large parts of the Russian wheat harvest. Thus, the challenges to food production are certain to continue; systems and network approaches will be invaluable tools for improving our understanding of the molecular workings of these organisms that we depend on so critically for our own existence.

SUMMARY POINTS

1. Systematic interactome studies allow specific biological questions to be answered beyond the development of hypotheses for individual proteins, and are necessary elements of systems biology.
2. Interaction assays are characterized by a high false-negative rate but give reliable data when implemented appropriately.
3. Interactome mapping requires stringent assay implementations to eliminate the accumulation of false positives.
4. Each interaction assay has a unique detection profile, and simple validation of a new interactome data set by an orthogonal assay requires benchmarking of the latter.
5. Benchmarking of assays using large sets of positive and negative control protein pairs gives estimates for the sensitivity and detection profile of a given assay. Use of standardized reference sets further increases the comparability and transparence of results.
6. Many bioinformatics resources are available to access and analyze interactome data.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Research in Unité de Recherche en Génomique Végétale is funded by the French National Institute for Agricultural Research (INRA), Evry University (UEVE), the National Center for Scientific Research (CNRS), and a grant from the 6th Framework Program of the European Commission (AGRONOMICS, LSHG-CT-2006-037704). J.V.L. is a postdoctoral fellow of the Research Foundation–Flanders.

LITERATURE CITED

1. Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature* 406:378–82
2. **Arabidopsis Interactome Mapp. Consort. 2011. Evidence for network evolution in an *Arabidopsis* interactome map. *Science* 333:601–7**
3. Backstrom S, Elfving N, Nilsson R, Wingsle G, Bjorklund S. 2007. Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. *Mol. Cell* 26:717–29

2. Describes the construction of AI-1, the first experimental proteome-scale interaction network for plants.

4. Book AJ, Gladman NP, Lee SS, Scalf M, Smith LM, Vierstra RD. 2010. Affinity purification of the *Arabidopsis* 26 S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* 285:25554–69
5. Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, et al. 2010. Functional modules in the *Arabidopsis* core cell cycle binary protein-protein interaction network. *Plant Cell* 22:1264–80
6. Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, et al. 2008. A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* 134:534–45
7. Bracha-Drori K, Shichrur K, Katz A, Oliva M, Angelovici R, et al. 2004. Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. *Plant J.* 40:419–27
8. Brandão MM, Dantas LL, Silva-Filho MC. 2009. AtPIN: *Arabidopsis thaliana* protein interaction network. *BMC Bioinforma.* 10:454
9. **Braun P. 2012. Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics* 12:1499–518**
10. **Braun P, Gingras AC. 2012. History of protein-protein interactions: from egg-white to complex networks. *Proteomics* 12:1478–98**
11. **Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, et al. 2009. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* 6:91–97**
12. Brizard JP, Carapito C, Delalande F, Van Dorsselaer A, Brugidou C. 2006. Proteome analysis of plant-virus interactome: comprehensive data for virus multiplication inside their hosts. *Mol. Cell. Proteomics* 5:2279–97
13. Brizuela L, Richardson A, Marsischky G, Labaer J. 2002. The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. *Arch. Med. Res.* 33:318–24
14. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* 10:2763–88
15. Cao S, Siriwardana CL, Kumimoto RW, Holt BF III. 2011. Construction of high quality Gateway entry libraries and their application to yeast two-hybrid for the monocot model plant *Brachypodium distachyon*. *BMC Biotechnol.* 11:53
16. Causier B. 2004. Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom. Rev.* 23:350–67
17. Causier B, Ashworth M, Guo W, Davies B. 2012. The TOPLESS interactome: a framework for gene repression in *Arabidopsis*. *Plant Physiol.* 158:423–38
18. Chang IF, Curran A, Woolsey R, Quilici D, Cushman JC, et al. 2009. Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in *Arabidopsis thaliana*. *Proteomics* 9:2967–85
19. Chen H, Zou Y, Shang Y, Lin H, Wang Y, et al. 2008. Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol.* 146:368–76
20. Chen J, Lalonde S, Obrdlik P, Noorani Vatani A, Parsa SA, et al. 2012. Uncovering *Arabidopsis* membrane protein interactome enriched in transporters using mating-based split ubiquitin assays and classification models. *Front. Plant Sci.* 3:124
21. Chen YC, Rajagopala SV, Stellberger T, Uetz P. 2010. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* 7:667–68
22. Ciannamea S, Kaufmann K, Frau M, Tonaco IA, Petersen K, et al. 2006. Protein interactions of MADS box transcription factors involved in flowering in *Lolium perenne*. *J. Exp. Bot.* 57:3419–31
23. Citovsky V, Gafni Y, Tzfira T. 2008. Localizing protein-protein interactions by bimolecular fluorescence complementation *in planta*. *Methods* 45:196–206
24. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2:2366–82
25. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, et al. 2009. Literature-curated protein interaction datasets. *Nat. Methods* 6:39–46
26. De Bodt S, Hollunder J, Nelissen H, Meulemeester N, Inze D. 2012. CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytol.* 195:707–20

9. Provides a comprehensive discussion of assay benchmarking and conclusions from it.

10. Gives a historical overview of protein interaction research from its beginnings to today's large-scale approaches.

11. Describes the comparative benchmarking of five different binary interaction assays using an identical set of 92 positive and 92 random protein pairs.

27. De Bodt S, Proost S, Vandepoele K, Rouze P, Van de Peer Y. 2009. Predicting protein-protein interactions in *Arabidopsis thaliana* through integration of orthology, gene ontology and co-expression. *BMC Genomics* 10:288
28. de Folter S, Immink RG, Kieffer M, Parenicova L, Henz SR, et al. 2005. Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell* 17:1424–33
29. Dreze M, Monachello D, Lurin C, Cusick ME, Hill DE, et al. 2010. High-quality binary interactome mapping. *Methods Enzymol.* 470:281–315
30. Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, 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
31. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, et al. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46:884–92
32. Ester C, Uetz P. 2008. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochem.* 9:29
33. Fields S, Song O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* 340:245–46
34. Finn RD, Marshall M, Bateman A. 2005. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 21:410–12
35. Fujikawa Y, Kato N. 2007. Split luciferase complementation assay to study protein-protein interactions in Arabidopsis protoplasts. *Plant J.* 52:185–95
36. Gadella TW Jr, van der Krogt GN, Bisseling T. 1999. GFP-based FRET microscopy in living plant cells. *Trends Plant Sci.* 4:287–91
37. Gassmann W, Bhattacharjee S. 2012. Effector-triggered immunity signaling: from gene-for-gene pathways to protein-protein interaction networks. *Mol. Plant-Microbe Interact.* 25:862–68
38. Gehl C, Waadt R, Kudla J, Mendel RR, Hansch R. 2009. New GATEWAY vectors for high throughput analyses of protein-protein interactions by bimolecular fluorescence complementation. *Mol. Plant* 2:1051–58
39. Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH, Geisler M. 2007. A predicted interactome for Arabidopsis. *Plant Physiol.* 145:317–29
40. Ghosh I, Hamilton A, Regan L. 2000. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* 122:5658–59
41. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. 2003. A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–36
42. Gong W, Shen YP, Ma LG, Pan Y, Du YL, et al. 2004. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol.* 135:773–82
43. Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, et al. 2009. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell* 138:314–27
44. Gu H, Zhu P, Jiao Y, Meng Y, Chen M. 2011. PRIN: a predicted rice interactome network. *BMC Bioinforma.* 12:161
45. He F, Zhou Y, Zhang Z. 2010. Deciphering the Arabidopsis floral transition process by integrating a protein-protein interaction network and gene expression data. *Plant Physiol.* 153:1492–505
46. Hu CD, Chinenov Y, Kerppola TK. 2002. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* 9:789–98
47. Hu CD, Kerppola TK. 2003. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* 21:539–45
48. Huang H, Bader JS. 2009. Precision and recall estimates for two-hybrid screens. *Bioinformatics* 25:372–78
49. Hunt AG, Xu R, Addepalli B, Rao S, Forbes KP, et al. 2008. Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. *BMC Genomics* 9:220
50. Immink RG, Ferrario S, Busscher-Lange J, Kooiker M, Busscher M, Angenent GC. 2003. Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genomics* 268:598–606
51. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 98:4569–74

39. Gives the first predicted *Arabidopsis* protein interactome based on the interolog method.

52. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, et al. 2000. Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* 97:1143–47
53. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, et al. 2003. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 302:449–53
54. Jensen LJ, Bork P. 2008. Not comparable, but complementary. *Science* 322:56–57
55. Karlova R, Boeren S, Russinova E, Aker J, Vervoort J, de Vries S. 2006. The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. *Plant Cell* 18:626–38
56. Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, et al. 2010. VirtualPlant: a software platform to support systems biology research. *Plant Physiol.* 152:500–15
57. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, et al. 2003. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA* 100:11394–99
58. Kerppola TK. 2008. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.* 37:465–87
59. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, et al. 2012. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 40:D841–46
60. Keskin O, Nussinov R, Gursoy A. 2008. PRISM: protein-protein interaction prediction by structural matching. *Methods Mol. Biol.* 484:505–21
61. Kim JG, Park D, Kim BC, Cho SW, Kim YT, et al. 2008. Predicting the interactome of *Xanthomonas oryzae* pathovar oryzae for target selection and DB service. *BMC Bioinforma.* 9:41
62. Kim Y, Min B, Yi GS. 2012. IDDI: integrated domain-domain interaction and protein interaction analysis system. *Proteome Sci.* 10(Suppl. 1):S9
63. Klopffleisch K, Phan N, Augustin K, Bayne RS, Booker KS, et al. 2011. Arabidopsis G-protein interactome reveals connections to cell wall carbohydrates and morphogenesis. *Mol. Syst. Biol.* 7:532
64. Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. 2011. Computational methods for Gene Orthology inference. *Brief. Bioinforma.* 12:379–91
65. Lalonde S, Sero A, Pratelli R, Pilot G, Chen J, et al. 2010. A membrane protein/signaling protein interaction network for *Arabidopsis* version AMPv2. *Front. Physiol.* 1:24
66. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al. 2012. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40:D1202–10
67. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY. 2010. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat. Biotechnol.* 28:149–56
68. Lee K, Thorneycroft D, Achuthan P, Hermjakob H, Ideker T. 2010. Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. *Plant Cell* 22:997–1005
69. Lee LY, Fang MJ, Kuang LY, Gelvin SB. 2008. Vectors for multi-color bimolecular fluorescence complementation to investigate protein-protein interactions in living plant cells. *Plant Methods* 4:24
70. Lensink MF, Wodak SJ. 2010. Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins* 78:3085–95
71. Leseberg CH, Eissler CL, Wang X, Johns MA, Duvall MR, Mao L. 2008. Interaction study of MADS-domain proteins in tomato. *J. Exp. Bot.* 59:2253–65
72. Li JF, Bush J, Xiong Y, Li L, McCormack M. 2011. Large-scale protein-protein interaction analysis in Arabidopsis mesophyll protoplasts by split firefly luciferase complementation. *PLoS ONE* 6:e27364
73. Li P, Zang W, Li Y, Xu F, Wang J, Shi T. 2011. AtPID: the overall hierarchical functional protein interaction network interface and analytic platform for *Arabidopsis*. *Nucleic Acids Res.* 39:D1130–33
74. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* 303:540–43
75. Li ZG, He F, Zhang Z, Peng YL. 2012. Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*. *Amino Acids* 42:2363–71
76. Liberman LM, Sozzani R, Benfey PN. 2012. Integrative systems biology: an attempt to describe a simple weed. *Curr. Opin. Plant Biol.* 15:162–67
77. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, et al. 2012. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40:D857–61

58. Provides a comprehensive and thorough discussion of BiFC methods and their applications.

78. Lievens S, Eyckerman S, Lemmens I, Tavernier J. 2010. Large-scale protein interactome mapping: strategies and opportunities. *Expert Rev. Proteomics* 7:679–90
79. Lin M, Shen X, Chen X. 2011. PAIR: the predicted *Arabidopsis* interactome resource. *Nucleic Acids Res.* 39:D1134–40
80. Lin M, Zhou X, Shen X, Mao C, Chen X. 2011. The predicted *Arabidopsis* interactome resource and network topology-based systems biology analyses. *Plant Cell* 23:911–22
81. Link AJ, Labaer J. 2008. Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 1: coating glass slides with amino silane. *Cold Spring Harb. Protoc.* 2008:pdb.prot5056
82. Liu C, Zhang J, Zhang N, Shan H, Su K, et al. 2010. Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. *Mol. Biol. Evol.* 27:1598–611
83. Luker KE, Piwnica-Worms D. 2004. Optimizing luciferase protein fragment complementation for bioluminescent imaging of protein-protein interactions in live cells and animals. *Methods Enzymol.* 385:349–60
84. Luker KE, Smith MC, Luker GD, Gammon ST, Piwnica-Worms H, Piwnica-Worms D. 2004. Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. USA* 101:12288–93
85. MacBeath G. 2002. Protein microarrays and proteomics. *Nat. Genet.* 32(Suppl.):526–32
86. Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, et al. 2005. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J. Am. Chem. Soc.* 127:146–57
87. Marcotte CJ, Marcotte EM. 2002. Predicting functional linkages from gene fusions with confidence. *Appl. Bioinforma.* 1:93–100
88. Maroun M, Aronheim A. 1999. A novel in vivo assay for the analysis of protein-protein interaction. *Nucleic Acids Res.* 27:e4
89. Martin K, Kopperud K, Chakrabarty R, Banerjee R, Brooks R, Goodin MM. 2009. Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *Plant J.* 59:150–62
90. Mendez-Rios J, Uetz P. 2010. Global approaches to study protein-protein interactions among viruses and hosts. *Future Microbiol.* 5:289–301
91. Miller JP, Lo RS, Ben-Hur A, Desmarais C, Stagljar I, et al. 2005. Large-scale identification of yeast integral membrane protein interactions. *Proc. Natl. Acad. Sci. USA* 102:12123–28
92. Morell M, Espargaro A, Aviles FX, Ventura S. 2008. Study and selection of *in vivo* protein interactions by coupling bimolecular fluorescence complementation and flow cytometry. *Nat. Protoc.* 3:22–33
93. Morsy M, Gouthu S, Orchard S, Thorneycroft D, Harper JF, et al. 2008. Charting plant interactomes: possibilities and challenges. *Trends Plant Sci.* 13:183–91
94. Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, et al. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333:596–601
95. Multinatl. Arabidopsis Steer. Comm. 2011. *The multinational coordinated* Arabidopsis thaliana *functional genomics project.* Annu. Rep., Multinatl. Arabidopsis Steer. Comm. **http://www.arabidopsis.org/portals/masc/2011_MASC_Report.pdf**
96. Nesvizhskii AI. 2012. Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* 12:1639–55
97. Nishimura MT, Dangl JL. 2010. Arabidopsis and the plant immune system. *Plant J.* 61:1053–66
98. Ohad N, Shichrur K, Yalovsky S. 2007. The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiol.* 145:1090–99
99. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, et al. 2012. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods* 9:345–50
100. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, et al. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38:D196–203
101. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, et al. 2010. NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* 464:788–91

94. Utilizes and expands on AI-1 to study how pathogens attack their hosts.

102. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* 96:4285–88
103. Petschnigg J, Wong V, Snider J, Stagljar I. 2012. Investigation of membrane protein interactions using the split-ubiquitin membrane yeast two-hybrid system. *Methods Mol. Biol.* 812:225–44
104. Plewczynski D, Ginalski K. 2009. The interactome: predicting the protein-protein interactions in cells. *Cell. Mol. Biol. Lett.* 14:1–22
105. Popescu SC, Popescu GV, Bachan S, Zhang Z, Seay M, et al. 2007. Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. *Proc. Natl. Acad. Sci. USA* 104:4730–35
106. Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, et al. 2009. MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev.* 23:80–92
107. Prieto C, De Las Rivas J. 2006. APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res.* 34:W298–302
108. Pusch S, Harashima H, Schnittger A. 2012. Identification of kinase substrates by bimolecular complementation assays. *Plant J.* 70:348–56
109. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, et al. 2004. Self-assembling protein microarrays. *Science* 305:86–90
110. Ransone LJ. 1995. Detection of protein-protein interactions by coimmunoprecipitation and dimerization. *Methods Enzymol.* 254:491–97
111. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, et al. 2003. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* 34:35–41
112. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, et al. 2005. Probabilistic model of the human protein-protein interaction network. *Nat. Biotechnol.* 23:951–59
113. Richardson LG, Howard AS, Khuu N, Gidda SK, McCartney A, et al. 2011. Protein-protein interaction network and subcellular localization of the *Arabidopsis thaliana* ESCRT machinery. *Front. Plant Sci.* 2:20
114. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17:1030–32
115. Rohila JS, Chen M, Cerny RL, Fromm ME. 2004. Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* 38:172–81
116. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, et al. 2006. Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* 46:1–13
117. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, et al. 2009. Protein-protein interactions of tandem affinity purified protein kinases from rice. *PLoS ONE* 4:e6685
118. Roux M, Schwessinger B, Albrecht C, Chinchilla D, Jones A, et al. 2011. The *Arabidopsis* leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell* 23:2440–55
119. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437:1173–78
120. Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, et al. 2005. An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. *Plant J.* 41:767–78
121. Ruokolainen S, Ng YP, Albert VA, Elomaa P, Teeri TH. 2010. Large scale interaction analysis predicts that the *Gerbera hybrida* floral E function is provided both by general and specialized proteins. *BMC Plant Biol.* 10:129
122. Schoonheim PJ, Veiga H, da Costa Pereira D, Friso G, van Wijk KJ, de Boer AH. 2007. A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. *Plant Physiol.* 143:670–83
123. Seo YS, Chern M, Bartley LE, Han M, Jung KH, et al. 2011. Towards establishment of a rice stress response interactome. *PLoS Genet.* 7:e1002020
124. Shekhawat SS, Ghosh I. 2011. Split-protein systems: beyond binary protein-protein interactions. *Curr. Opin. Chem. Biol.* 15:789–97
125. Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmens I, et al. 2009. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat. Methods* 6:47–54

126. Simonsen M, Maetschke SR, Ragan MA. 2012. Automatic selection of reference taxa for protein-protein interaction prediction with phylogenetic profiling. *Bioinformatics* 28:851–57

127. One of the first plant studies to combine GFP-based AP-MS with label-free identification to discriminate true and false positives.

127. Smaczniak C, Immink RG, Muino JM, Blanvillain R, Busscher M, et al. 2012. Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. USA* 109:1560–65

128. Smith GR, Sternberg MJ. 2002. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* 12:28–35

129. Snider J, Kittanakom S, Curak J, Stagljar I. 2010. Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. *J. Vis. Exp.* 36:e1698

130. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. 2011. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* 39:D698–704

131. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122:957–68

132. Subramaniam R, Desveaux D, Spickler C, Michnick SW, Brisson N. 2001. Direct visualization of protein interactions in plant cells. *Nat. Biotechnol.* 19:769–72

133. Suderman M, Hallett M. 2007. Tools for visually exploring biological networks. *Bioinformatics* 23:2651–59

134. Swatek KN, Graham K, Agrawal GK, Thelen JJ. 2011. The 14-3-3 isoforms chi and epsilon differentially bind client proteins from developing *Arabidopsis* seed. *J. Proteome Res.* 10:4076–87

135. Provides a reference integrative database focused on experimental and predicted protein-protein interactions.

135. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, et al. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39:D561–68

136. Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I. 2010. Identification and characterization of nuclear pore complex components in *Arabidopsis thaliana*. *Plant Cell* 22:4084–97

137. Tsesmetzis N, Couchman M, Higgins J, Smith A, Doonan JH, et al. 2008. *Arabidopsis* reactome: a foundation knowledgebase for plant systems biology. *Plant Cell* 20:1426–36

138. Tzfira T, Vaidya M, Citovsky V. 2004. Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. *Nature* 431:87–92

139. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–27

140. Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L. 2011. A kaleidoscopic view of the *Arabidopsis* core cell cycle interactome. *Trends Plant Sci.* 16:141–50

141. Van Leene J, Eeckhout D, Persiau G, Van De Slijke E, Geerinck J, et al. 2011. Isolation of transcription factor complexes from *Arabidopsis* cell suspension cultures by tandem affinity purification. *Methods Mol. Biol.* 754:195–218

142. Demonstrates the superiority of the GS tag for TAP in plants.

142. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, et al. 2010. Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 6:397

143. Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, et al. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 6:1226–38

144. Van Leene J, Witters E, Inze D, De Jaeger G. 2008. Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci.* 13:517–20

145. Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, et al. 2012. PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* 40:D445–52

146. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, et al. 2009. An empirical framework for binary interactome mapping. *Nat. Methods* 6:83–90

147. Vidal M, Cusick ME, Barabasi AL. 2011. Interactome networks and human disease. *Cell* 144:986–98

148. Waadt R, Schmidt LK, Lohse M, Hashimoto K, Bock R, Kudla J. 2008. Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes *in planta*. *Plant J.* 56:505–16

149. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, et al. 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287:116–22

150. Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, et al. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* 40:428–38
151. Weiste C, Iven T, Fischer U, Onate-Sanchez L, Droge-Laser W. 2007. *In planta* ORFeome analysis by large-scale over-expression of GATEWAY-compatible cDNA clones: screening of ERF transcription factors involved in abiotic stress defense. *Plant J.* 52:382–90
152. White MA. 1996. The yeast two-hybrid system: forward and reverse. *Proc. Natl. Acad. Sci. USA* 93:10001–3
153. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. 2000. DIP: the database of interacting proteins. *Nucleic Acids Res.* 28:289–91
154. Xu F, Li G, Zhao C, Li Y, Li P, et al. 2010. Global protein interactome exploration through mining genome-scale data in *Arabidopsis thaliana*. *BMC Genomics* 11(Suppl. 2):S2
155. Xu X, Soutto M, Xie Q, Servick S, Subramanian C, et al. 2007. Imaging protein interactions with bioluminescence resonance energy transfer (BRET) in plant and mammalian cells and tissues. *Proc. Natl. Acad. Sci. USA* 104:10264–69
156. Yamada K, Lim J, Dale JM, Chen H, Shinn P, et al. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302:842–46
157. Yang PZ, Fu HY, Walker J, Papa CM, Smalle J, et al. 2004. Purification of the *Arabidopsis* 26 S proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J. Biol. Chem.* 279:6401–13
158. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, et al. 2008. High-quality binary protein interaction map of the yeast interactome network. *Science* 322:104–10
159. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, et al. 2011. Next-generation sequencing to generate interactome datasets. *Nat. Methods* 8:478–80
160. Yu X, Ivanic J, Memisevic V, Wallqvist A, Reifman J. 2011. Categorizing biases in high-confidence high-throughput protein-protein interaction data sets. *Mol. Cell. Proteomics* 10:M111.012500
161. Zhong J, Haynes PA, Zhang S, Yang X, Andon NL, et al. 2003. Development of a system for the study of protein-protein interactions *in planta*: characterization of a TATA-box binding protein complex in *Oryza sativa*. *J. Proteome Res.* 2:514–22
162. Zhu P, Gu H, Jiao Y, Huang D, Chen M. 2011. Computational identification of protein-protein interactions in rice based on the predicted rice interactome network. *Genomics Proteomics Bioinforma.* 9:128–37

Annual Review of Plant Biology
Volume 64, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Plant Biology* articles may be found at http://www.annualreviews.org/errata/arplant