UNIVERSITEIT GENT

FACULTEIT WETENSCHAPPEN

Department of Molecular Genetics

Study of the cell cycle interactome of *Arabidopsis thaliana* through a targeted proteomics approach

Jelle Van Leene

Academic year: 2008-2009

Promoter: Prof. Dr. Geert De Jaeger

Co-promoter: Prof. Dr. Dirk Inzé

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor in Science: Biotechnology

Plant Systems Biology

VIB

***"No pain, no gain"* by Robert Herrick**

Onderzoek gefinancierd met een specialisatiebeurs van het Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen) (Generisch Basisonderzoek aan de Universiteiten [GBOU] beurs nummer 20193)

Research funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) (Generisch Basisonderzoek aan de Universiteiten [GBOU] grant number 20193)

Dankwoord

Het heeft bloed, zweet en tranen (zweet en tranen van te lachen dan vooral) gekost, maar na meer dan 5 jaar is het einde van de tunnel eindelijk in zicht. Het moment is dan ook aangebroken om iedereen te bedanken die me heeft bijgestaan en me heeft geholpen bij het behalen van mijn doctoraat.

Ik herinner me het nog alsof het gisteren was: ik zat op de trein, waarschijnlijk weer aan het flirten met 1 of ander meisje, toen ik plots telefoon kreeg. Het was Geertje. Wie is Geertje? Geertje kunnen we beschouwen als den Jean-Marie Pfaff van Oost-Vlaanderen, hij is namelijk den besten keeper van Wippelgem en omstreken, en dat wil wat zeggen hoor. Meer dan dat nog, was hij vooral de persoon die me tijdens mijn licentiaatsthesis heeft begeleid en me de kans heeft gegeven om aan wetenschap te doen. Hoewel we verbannen waren tot de containers van de Ledeganck, kwam ik daar in een zeer aangename groep terecht en leerde ik er ook Dominique e.a. kennen. Nu, om 1 of andere obscure reden - misschien was het wel omdat Geert en ik meester zijn in het o zo verfijnde meetjeslandse dialect, wat ons onderscheidde van de meeste andere wetenschappers - moet Geert toch vertrouwen in mij gehad hebben, want hij had een TAP project binnen gehaald, wist hij me te vertellen. Ja, had ik dat goed gehoord?? Een TAP project, tappen…daar moest ik wel op in gaan natuurlijk. Geertje, bedankt voor het vertrouwen en alles dat je me hebt bijgeleed.

Als een klein kind dat naar zijn eerste schooldag gaat, begaf ik me op de 1ste juli 2003 - uitgerekend de 1ste dag van de grote vakantie (snik snik) - naar het nieuwe gebouw in Zwijnaarde. Daar gekomen nam ik kennis van mijn nieuwe collega'tjes: Dominique, Hilde, Annelies en Geert Persiau (vanaf nu Geper genoemd) had ik al eerder ontmoet, dus dat ging wel goed komen. Zeker het zien van den dienen met dat sikske en met de piercings stelde me gerust, want ik vond bij hem iemand van mijn eigen soortgenoten – de primatus marginalis - terug. Maar wie waren die 2 nieuwelingetjes daar? Oeioei, en ze hadden samen gestudeerd, hopelijk gingen ze geen samenzwering beginnen tegen mij. Het waren Eveline en Gert. Samen begonnen we aan onze eerste werkdag, kasten kuisen, my favorite job, dus

dat was al goed begonnen. Bij Eveline vond ik onmiddellijk alles wat ik zocht, iemand bij wie ik de volgende 5 jaar altijd terecht kon en die O zo goed voor me zorgde, en ja, ik ben nogal een mama's kindje en had daar dus behoefte aan. Na de deskundige klonerings expertise die Dominique Eveline en mij had aangeleerd, sloeg Eveline er dadelijk in alle nodige klonen aan te maken (voor de niet-wetenschapper onder de lezers: wie problemen heeft met klonen, je weet bij wie je terecht moet). In Gert zag ik de persoon die ik in feite wou zijn, hij heeft altijd alles zo goed voor elkaar, hij komt met iedereen overeen, je kan alles aan hem kwijt - bij anderen ook hoor, maar bij hem wist je zeker dat het in vertrouwen was -, niets is hem teveel, hij is een uitstekende wetenschapper, en ondanks hij van dezelfde prille leeftijd is als ik, heeft hij het toch voor elkaar gespeeld om zich te laten omringen door een prachtig vrouwtje Sarah en 2 schatten van kindjes. Maar mam en paps, wees gerust, op een dag zal jullie geduld ook beloond worden hoor. Samen met Eveline besloot Gert de culturele kant op te gaan (of werden ze hiertoe gedwongen?), en elke woensdag waren ze voortaan zij aan zij te vinden om onze cultuurkes in leven te houden.

Ons Hilde was samen met Geper afkomstig uit de groep geleid door de nar Peter Casteels, en zij vormden beiden de onmisbare link die we misten met de proteinekes. Zonder hun eiwitexpertise stonden we nu 5 jaar later waarschijnlijk nog altijd nergens. Daar Geertje ondertussen gepromoveerd was tot professor - onderschat die mannekes van het meetjesland dus niet -, was er nood aan iemand als Hilde, die de touwtjes overnam om alles te coördineren en in goede banen te leiden, en ze heeft dit met verve uitgevoerd. Geper, oneindig maal dank voor alle TAPkes die je uitgevoerd hebt, ik heb je op geen enkel moment horen klagen, en alle zuiveringen werden met een immense precisie en fingerspitchengefühl uitgevoerd, een echt huzarenstukje!

Het grootste struikelblok in dit project ging ongetwijfeld de identificatie van de eiwitten worden, een uitdaging die superwoman Dominique voor haar rekening nam. Daar de technologie voor deze analyse niet echt voorhanden was bij ons, werd er gezocht naar een vruchtvolle samenwerking die we uiteindelijk vonden bij de mannen van 't stad, in Antwerpen. Elke dag opnieuw doorstond ons Dominique de eindeloze files tussen Gent en Antwerpen, maar tot 1 ieders verbazing had Dominique daar geen last van. Hmmm, wat zat daar achter, haar liefde voor de wetenschap, of toch iets anders? Al gauw bleek dat ze het zoeken naar interacties iets te letterlijk had opgenomen, en binnen de kortste keren stampte

ze samen met Erwin en de andere leden van Ceproma een MS-platform uit de grond dat kon wedijveren met de rest van de wereld. Natuurlijk werden zij op hun beurt gesteund door anderen zoals daar zijn Noor, Anne, Kim en Kris, die allen onmisbare schakels in het netwerk bleken te zijn. Noortje, nogmaals bedankt voor de onvergetelijke reis naar Maratea, het was daar dikwijls afzien in dat paradijs aan het strand, maar toch bedankt dat je mijn klamme handje hebt vast gehouden, zowel op de vlieger als net voor mijn presentatie.

Zoals sommige bizarre resultaten al deden vermoeden - even dacht ik al dat we een science paper binnen hadden, daar fertilisatie in planten op een bepaald moment zeer analoog bleek te zijn met die bij mensen - zou het ook een zeer vruchtbare periode worden. Nogmaals proficiat aan Geertje, Gert, Hilde en Noor met jullie prachtige kindjes.

Nu alles op wieltjes liep, gebeurden er een paar verschuivingen, sommigen keerden terug naar hun oorspronkelijke 'roots', anderen zochten warmere oorden op in Kreta of vonden terug onderdak in het kasteel van Peter. Dat opende natuurlijk de deuren voor anderen en van heide en verre stroomden ze toe. De intelligentie coëfficiënt van de groep schoot plots de hoogte in met de komst van stand-up comedian Jan, en mijn allerliefste Aurine'tje. Steevast klom Aurine'tje naar de top, en vaak liet ze mij het kijken na, wat ik uiteraard helemaal niet erg vond. Hoewel Jan er nog altijd niet in geslaagd is om een perfecte imitatie van mij neer te zetten - dat meetjeslands dialect is toch zo ondoorgrondbaar - bracht hij me heel wat inzicht bij en hervond ik mijn liefde voor moedertje Natuur. Vervolgens werd het meetjeslands clubje verder uitgebreid met de komst van Astrid, we waren nu een onoverwinnelijk blok aan het worden. De volgende twee die in het rijtje stonden om bij ons te komen waren Yelle, die me soms met verstomming doen kon verbazen, maar nu als een volleerd meester met engelengeduld onze gellekes – of zjeelekes voor de mensen van Antwerpen – behandelt, en…trommelgeroffel…(en getetter)… ons Sandy'tje, nog een soortgenootje van mij en Geper. Even vlug als ze in mijn leven gekomen is, is ze nu weeral verdwenen, maar ik weet dat ik aan haar een goed vriendinnetje bij heb, en onze paden zullen zich ongetwijfeld nog kruisen. Ook danke schön aan Jens, Stefanie, Steven en Klaas (ook Kris past hier weer in het rijtje) voor mij in te leiden tot de skills van de bioinformatics.

Natuurlijk was niks van dit wondermooie verhaal mogelijk zonder de steun en het vertrouwen van Dirk en zijn entourage, van Hilde, Rebecca, Mansour, Jacky, Karel, Nancy, Martine, de mensen van IT, …en allen die de echte peilers vormen van dit departement.

Thanks aan mijn o zo gehoorzame studentjes Guido en Emilie!

Dank aan al mijn maatjes en vriendinnekes, teveel om op te noemen, maar toch special thanks voor m'n grote zwak, de ladies, namelijk Sofie'tje, Maya'ke, Katrientje, Talia'tje en uiteraard mijn rots in de branding, Liesio!!

Bedankt aan allen die ik vergeten ben, maar die ongetwijfeld in mijn hart zitten.

Tenslotte gaat mijn allergrootste dank uit naar mijn familie, naar mijn mammie en pappie en mijn liefste broerke Bart, bedankt voor jullie onvoorwaardelijke steun, bedankt dat ik altijd mijn eigen weg mocht inslaan, bedankt voor alles!!

Jelle,

01 december '08

Examination Committee

Promoter: Prof. Dr. Geert De Jaeger	VIB-department of Plant Systems Biology
(Secretary)	Department of Molecular Genetics
	Ghent University, Belgium
Co-promoter: Prof. Dr. Dirk Inzé	VIB-department of Plant Systems Biology
	Department of Molecular Genetics
	Ghent University, Belgium
Chairman: Prof. Dr. Anna Depicker	VIB-department of Plant Systems Biology
	Department of Molecular Genetics
	Ghent University, Belgium
Dr. Tilmann Bürckstümmer	Ce-MM, Research Centrum for Molecular Medicine
	Austrian Academy of Sciences, Austria
Prof. Dr. Bart Devreese	Laboratory for Protein Biochemistry and Protein Engineering
	Department of Biochemistry, Physiology and Microbiology
	Ghent University, Belgium
Dr. Pierre Hilson	VIB-department of Plant Systems Biology
	Department of Molecular Genetics
	Ghent University, Belgium
Prof. Dr. Lieven De Veylder	VIB-department of Plant Systems Biology
	Department of Molecular Genetics
	Ghent University, Belgium
Dr. Erwin Witters	Department of Chemistry
	University of Antwerp, Belgium
Prof. Dr. Harry Van Onckelen	Department of Biology
	University of Antwerp, Belgium
Prof. Dr. Rudy Beyaert	VIB-department of Molecular Biomedical Research
	Department of Molecular Biology
	Ghent University, Belgium

Table of Contents

List of abbreviations

2-DGE	two-dimensional gel electrophoresis
APC	anaphase-promoting complex
AQUA	absolute protein quantifications
BiFC	bimolecular fluorescence complementation
BN-PAGE	blue native PAGE
BRET	bioluminescence resonance energy transfer
CAK	CDK-activating kinase
CBP	calmodulin-binding peptide
CDK	cyclin dependent kinase
CFP	cyan fluorescent protein
COFRADIC	combined fractional diagonal chromatography
Co-IP	co-immunoprecipitation
CTD	C-terminal domain
CYC	cyclin
DIGE	difference gel electrophoresis
dSLAM	diploid-based synthetic-lethality analysis on microarrays
ESI	electron spray ionization
EST	expressed sequence tag
FACS	Fluorescence-activated cell sorting
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
GO	gene ontology
GST	glutathione S-transferase
GUS	β-glucuronidase
HA	hemagglutinin
HILEP	hydroponic isotope labeling of entire plants
HPLC	high pressure liquid chromatography
ICAT	isotope-coded affinity tags
IgG	immunoglobulin G
IMAC	immobilized metal affinity chromatography
IP	immunoprecipitation

iTRAQ	isobaric tag for relative and absolute quantification
KRP	KIP-related protein
LC	liquid chromatography
MALDI	matrix-assisted laser desorption ionization
MBP	maltose-binding protein
MS	mass spectrometry
MSA	M-specific activator
MudPIT	multidimensional protein identification technology
NLS	nuclear localization signal
ORF	open-reading-frame
PCA	protein-fragment complementation assay
PCC	pearson correlation coefficient
PMF	peptide mass fingerprinting
PPI	protein-protein interaction
PTM	post-translational modification
SBP	streptavidin binding peptide
SILAC	stable isotope-labeled amino acid in cell culture
SILIP	stable isotope labeling *in planta*
SMR	SIAMESE-Related
TAP	tandem affinity purification
TEV	tobacco etch virus
TF	transcription factor
TOF	time-of-flight
Y2H	yeast two-hybrid
YFP	yellow fluorescent protein

Preface

To understand how genomes are translated into a living organism, like the model plant *Arabidopsis thaliana*, it is essential to examine not only the DNA sequence and the genes expressed from this sequence, but also the proteins, i.e. the translated products of the genes. Since proteins execute and control the majority of cellular processes, they reflect more directly the cellular biochemistry than genes do.

In chapter 1, an overview with a focus on plant studies is given of the most important technologies covered by proteomics, enabling a system-wide analysis of proteins. These methods analyze different aspects of proteins, such as their quantitative expression, localization, structure, and state of modification. However, the major emphasis of the introduction lies in the analysis of protein-protein interactions, as this is the topic of the experimental work presented in this thesis. Interactions between proteins play a critical role in the vast majority of cellular processes, and their analysis allows further functional annotation of genes. To date, the most widely used methods to analyze protein-protein interactions are the Yeast Two-Hybrid method, mapping direct binary protein interactions, and methods based on isolation of protein complexes by affinity purification and protein identification through mass spectrometry.

As one-step affinity purification suffers from low specificity, a more elegant approach called tandem affinity purification (TAP) was developed in *Saccharomyces cerevisiae*, residing on two consecutive affinity chromatography steps based on the fusion of a bait protein to a dual affinity tag. As described in chapter 2, we developed a TAP-based technology platform to isolate and characterize protein complexes from suspension-cultured plant cells. This platform covers cloning of transgenes encoding TAP tagged bait proteins, transformation of these transgenes in *Arabidopsis thaliana* cell suspension cultures, protein extract preparation from these plant cells, tandem affinity purification of the bait protein with its interacting molecules, and protein identification by mass spectrometry. With the final aim to built a cell cycle interactome, fast growing cell suspension cultures were chosen as this is an ideal system to study protein complexes involved in cell division. As proof of concept, results are presented that were obtained with 6 cell cycle bait proteins.

Despite the successful transformation of the yeast TAP method to *Arabidopsis*, some problems are still associated with the technology when applying plant protein extracts, as illustrated by the low number of TAP purified complexes reported so far. The traditional TAP method still faces limitations as low specificity and low protein complex yield. So to further optimize the method and to bring protein complex analysis from plants tissues to its full potential, we were continuously in search for optimized versions of the TAP tag. Improved results concerning specificity and bait protein accumulation levels were obtained with two alternative TAP tags, namely the GS tag, developed for TAP in mammalian cells, and the in-house developed CSFH tag, providing alternatives when one of the two tags fails. This alternative TAP tag evaluation screen is discussed in chapter 3 and in the supplement accompanying chapter 3.

Finally, as discussed in chapter 4, this TAP-based technology platform was used to map the cell cycle interactome of *Arabidopsis thaliana*. Protein complexes involving 102 proteins related to cell cycle were isolated and characterized, delineating 857 interactions among 393 proteins. The quality of the cell cycle interactome is assessed through an integrative approach combining transcript co-expression values, gene ontology similarities, and cell cycle-related features. Biological important gene networks were extracted from the interactome and described. The interactome may serve as a hypothesis-generating tool to further extend our knowledge of cell division and plant growth and development.

Chapter 1

Functional proteomics in plants

1.1 Introduction

Plants form the basis for the survival of all higher organisms on Earth. A systematic molecular evaluation of the complete genetic information and the resulting cellular activities will be essential to further unravel the biology of plants and will greatly impact our means of utilizing plants for the benefits of humanity and environment.

A major landmark in plant research has been the availability of complete genome sequences, such as for the model plant *Arabidopsis thaliana* (mouse-ear cress), for *Oryza sativa* (rice), *Populus trichocarpa* (black cottonwood), *Medicago truncatula* (barrel medic) and of enormous EST databases for a wide variety of other plant species (tobacco, tomato, maize, potato, wheat, soybean, lotus, etc…). This allowed a paradigm shift from the 'one gene – one hypothesis' approach to more global, systematic strategies that analyze genes or proteins on a genome- and proteome-wide scale.

As the blueprints of several plants became available, genome annotation was done and a first level of functional annotation was achieved by homology searches with sequences from other, better characterized organisms (Vandepoele et al., 2002; Capron et al., 2003a; Shultz et al., 2007). However, genome sequence *per se* is not sufficient to explain and predict cellular phenomena, as it is largely the proteins that execute and control the majority of cellular processes, i.e. DNA replication and transcription, progression through the cell cycle, protein synthesis and degradation, regulation of metabolic and signaling pathways, as well as a myriad of minor but important functions. It are the proteins that form the bridge between genes and the phenotype, and proteomics is intricately linked to allied –omics including genomics, transcriptomics, and metabolomics.

The term proteomics was introduced in 1994 at the *Conference on Genome and Protein Maps* (Siena, Italy) as the 'PROTEin complement expressed by a genOME' (Wasinger et al., 1995). Thus a proteome study is expected to represent a comprehensive survey of all proteins expressed at a specific time point, under certain conditions, in a given tissue. Furthermore, in addition to their primary amino-acid sequence, other properties of proteins such as their state of modification and association with other proteins or molecules of different types, their relative amounts, specific activity, subcellular localization, and three-

dimensional structure, represent crucial information to understand the function of a protein and ultimately, for the description of complex biological systems (Figure 1). So, the term proteomics is no more restricted to the construction of protein repertoires, but is now extended to studies of protein properties like protein-protein interactions and post-translational modifications, the field of 'functional' proteomics.

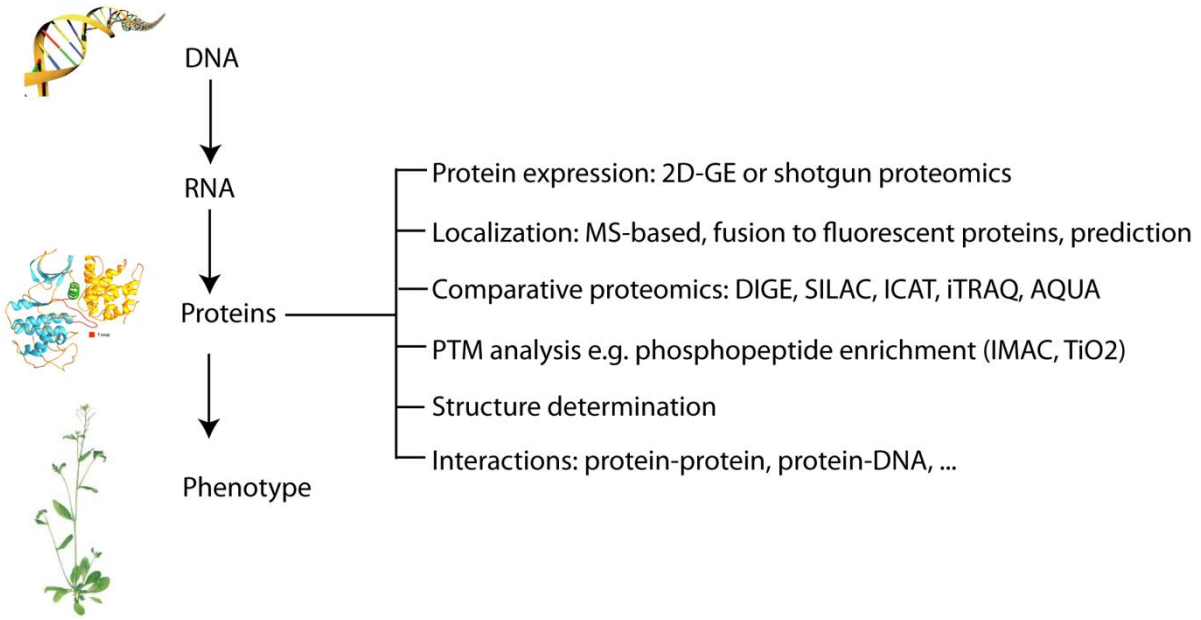


Figure 1: Proteomics analyzes different aspects of proteins: which proteins are expressed, where and to what extend they are expressed, how they look like (structure + post-translational modifications (PTMs)), and last but not least, with which other molecules (proteins, DNA, RNA, lipids, metabolites) they associate.

So far, for approximately 26 % of the *Arabidopsis* genes the function remains unknown according to the functional Gene Ontology (GO) category listed by the *Arabidopsis* Information Resource (TAIR) (Rhee et al., 2003). Of the remainder, a huge proportion lack complete or adequate functional annotation, demonstrating the need for systematic and large scale studies at the protein level. However, proteome studies face huge challenges: there is the high degree of structural and physicochemical heterogeneity of proteins, which is directly linked to the diversity of their functions within a cell, making the analysis of proteins much more technically demanding than genome or transcriptome analyses. Protein expression levels within a cell vary several orders of magnitude. Rare proteins are present in the order of 10-100 copies per cell, whereas the most abundant proteins are present at levels between 10^5 and 10^7 molecules per cell, meaning that the few most proteins, e.g.

rubisco in leaves and storage proteins in seeds often interfere with detection of low level proteins. There is no technology to amplify low abundant proteins, so the amount of starting tissue and detection sensitivity are critical limitations. Furthermore, while the *Arabidopsis* genome contains approximately 27235 protein coding genes according to the TAIR8.0 release, the corresponding proteome is much more complex. Events such as alternative transcription initiation and alternative splicing of mRNAs, alternative initiation of translation, and post-translational modifications (PTM's) generate a highly diverse set of proteins that could exceed a million distinct molecular species within a given cell. About 453 potential protein modifications have been reported (Garavelli, 2004), and a protein may contain multiple different PTMs at any given time. In addition, a proteome is dynamic: it changes during development and in response to external stimuli, and the proteins assemble into protein complexes forming interaction and regulatory networks. Finally, another level of complexity is added when we look at protein complexes as entities of biological activity that serve to create functional diversity by contextual combination of gene products. So, to fully understand the cellular machinery or even a single biological function, simply listing the proteins is not enough, all the interactions between them need to be delineated as well, PTM's need to be mapped, and quantitative profiling of proteins is required.

1.2 Protein expression and localization

1.2.1 Mapping the proteins

The last decade, DNA chips have enabled gene expression analysis through the comprehensive analysis of transcript levels. This was made possible mainly by the chemically homogeneous character of RNA and its easy extraction, amplification and sequencing. The *Arabidopsis* genome was the first eukaryotic genome that was entirely represented on TILING arrays (Mockler et al., 2005). Inferred from microarrays, high-throughput expression data is stored in databases such as Genevestigator, and these allow plant biologists to explore biological processes (Zimmermann et al., 2005). Although variance in protein abundance can mostly be explained by mRNA abundance, due to post-transcriptional regulation, protein expression does not always correlate with mRNA levels, with Pearson correlation coefficients ranging from 0.46 to 0.76, depending on the technology used to

quantify mRNA and protein levels and on the organism (Anderson and Seilhamer, 1997; Hack, 2004; Lu et al., 2007). Furthermore, the proteome reflects the expression of the molecules that more directly influence cellular biochemistry; this provides a more accurate representation of cellular state than profiling the expression of mRNAs. Therefore, one of the objectives of proteomics is the documentation of as many as possible proteins. Although technologies as Fluorescence-activated cell sorting (FACS) allow selective isolation of single cell types and analysis of their transcriptome (De Smet et al., 2008), entire proteomes of single cell types have not yet been fully mapped, mainly because of limitations in sensitivity. A related strategy is to target subcellular proteomes, thereby dramatically reducing the protein complexity of a particular extract and revealing important information about subcellular localization. An extension of the theme is 'comparative' proteomics, with the aim to characterize differences between protein populations from different sources, e.g. wild type *versus* mutant plants, or tissues at different developmental stages or under different environmental conditions, rather than to identify all proteins of a particular biological sample.

Different methods are currently available for profiling protein expression (Agrawal et al., 2005), including two-dimensional gel electrophoresis (2-DGE) or liquid chromatography (LC) coupled to mass spectrometry (MS). Standard protein profiling technologies have two major shortcomings: due to the high degree of heterogeneity of proteins, it is not possible to extract and solubilize the entire proteome of a tissue with a single protein extraction protocol, especially due to the difficulties associated with the solubilization of hydrophobic membrane-associated proteins. Second, the dynamic range of protein concentrations in a cell exceeds the sensitivity of most mass spectrometry devices, and as a consequence only the most abundant proteins are identified. Next to these general drawbacks, plant tissues generate additional problems for proteomic analyses (Rose et al., 2004; Isaacson et al., 2006). There are many plant tissue compounds from the secondary metabolism which negatively affect protein extraction, separation and subsequent analysis (Tsugita and Kamo, 1999). Differentiated plant cells have low protein concentrations due to their large vacuoles and rigid cell walls. Moreover, plant cells are rich in proteases, which require specific

precautions and elaborate protease inhibitor cocktails during extraction (Hochholdinger et al., 2006).

During 2-DGE complex protein mixtures are separated by molecular charge (pI) in the first dimension and by molecular mass in the second dimension. Next, proteins are visualized, spots are excised and proteins are identified by MS. Prerequisite for high-throughput protein identification is the availability of a fully sequenced and annotated genome. Although 2-DGE-based proteomics has proven powerful for the global analysis of proteins, it still retains technical problems that need to be solved (Corthals et al., 2000; Gygi et al., 2000). It is costly, labor- and time-consuming, and it has low gel-to-gel reproducibility, limiting high-throughput analysis of protein expression. Furthermore, problems arise for proteins with unusual low or high molecular weight, low or high pI, low abundance or high hydrophobicity. In addition, the entire protein profiling and quantification are not possible due to the limited loading capacity and incomplete staining methods (Park, 2004). Subcellular fractionation can be the solution, as this approach can dramatically reduce the complexity of protein extracts, while rare proteins are enriched and thus more readily to detect.

To overcome problems associated with 2-DGE, a gel-free approach named multidimensional protein identification technology (MudPIT) was developed (Washburn et al., 2001; Chen et al., 2006). Peptide mixtures generated from complex protein samples are first subjected to cation exchange and reverse phase HPLC and coupled in real-time for analysis by ESI-MS/MS. Often referred to as ‘shotgun proteomics’, gel-free methodologies based on LC-MS/MS reduce the complexity of peptide mixtures, leading to an increased proteome coverage. Reduction of the complexity of the peptide sample can also be achieved with combined fractional diagonal chromatography (COFRADIC) (Gevaert et al., 2003), allowing selective isolation of N-terminal peptides.

The last decade, different plant research groups focused on a variety of organelles and subcellular localizations and nuclear, chloroplast, mitochondrial, amyloplast, vacuolar, peroxisome, plasma membrane, endoplasmic reticulum, and cell wall subproteomes were exposed by either 2-DGE or LC-MS/MS approaches. Most of these proteomic studies are excellently reviewed (Canovas et al., 2004; Park, 2004; Baginsky and Gruissem, 2006;

Rossignol et al., 2006; Jorrin et al., 2007). Recently, a comprehensive proteome map was assembled for *Arabidopsis thaliana* from high-density, organ-specific proteome catalogues generated for different organs, developmental stages, and undifferentiated cultured cells. This map provides information about genome activity and proteome assembly and is available as a resource for plant systems biology (Baerenfaller et al., 2008).

1.2.2 Subcellular localization

Large-scale subcellular proteomics can give insight into the function of a protein or the compartment as a whole, but it can also provide information on the mechanism of protein targeting and trafficking (Chen and Harmon, 2006). An alternative approach to mass spectrometry to determine subcellular localizations is the expression and visualization of fluorescent proteins fused to a protein of interest (Tian et al., 2004; Koroleva et al., 2005; Li et al., 2006b). Next to this, the experimental localization data can be used to evaluate and improve current computer-based algorithms for signal peptide and intracellular targeting prediction based on primary amino-acid sequence. For *Arabidopsis*, localization information coming from both experimental data and *in silico* predictions is stored in a database called SUBA, the *Arabidopsis* subcellular database (Heazlewood et al., 2005; Heazlewood et al., 2007). In addition, marker proteins, labeling a specific organelle, are described for *Arabidopsis* and can be useful to determine the localization of unknown proteins (Nelson et al., 2007).

1.3 Protein quantification and differential expression profiling

Changes in gene expression are usually monitored at transcript level by quantitative real time reverse transcription-PCR (Nolan et al., 2006) or genome-wide with DNA chips. Maps of co-expressed genes can be inferred from these microarray data stored in specific databases through web-based analysis tools such as those provided by ATTED-II (Obayashi et al., 2007) and others (Steinhauser et al., 2004; Toufighi et al., 2005; Manfield et al., 2006; Srinivasasainagendra et al., 2008). As mentioned before however, survey of protein accumulation levels can offer complementary information besides transcript levels. Moreover, it was recently demonstrated that proteomics and transcriptomics are becoming equally comprehensive (Cox and Mann, 2007).

The aim of differential protein expression profiling is to identify protein level fluctuations between different biological samples: one may be interested in proteins specific for certain tissues or organs, or in proteins that are up- or down-regulated during different growth and plant developmental stages, upon hormone treatment, by disease and stress (low temperature, heat, drought, salt, ozone, …) or in different genotypes (Hirano et al., 2004). Here, comparative proteomic analyses are indispensable to study protein amount dynamics.

Until recently, the only possibility to study protein dynamics was the use of comparative 2-DGE, during which spot patterns and intensities were compared using sophisticated image analysis software. Typical problems arise due to gel-to-gel variations, leading to high false-positive and false-negative rates. These limitations may be overcome by analyzing different samples in the same gel, a technology termed DIGE, for difference gel electrophoresis (Figure 2). In DIGE, proteins from two different samples are labeled separately with one of the two fluorescent dyes Cy3 or Cy5. These two samples are then combined with a third mixture containing equal amounts of the two samples, labeled with Cy2 for internal calibration (Tonge et al., 2001; Karp et al., 2004; Casati et al., 2005). After separation by 2-DGE, changes in protein amounts are visualized by scanning the gel with different lasers and overlapping the images. This approach still faces most of the classical problems associated with 2-DGE, discussed above.

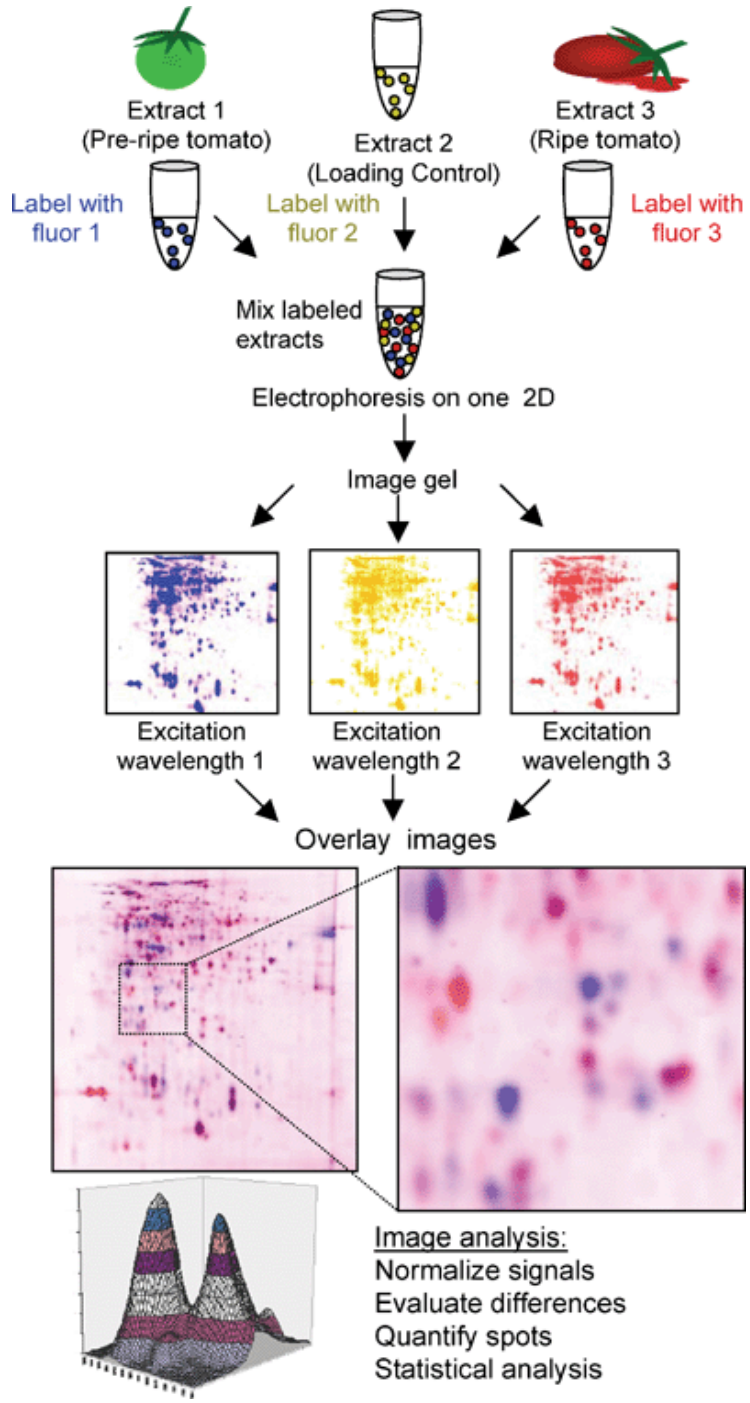


Figure 2: Outline of the difference in-gel electrophoresis (DIGE) approach for comparative proteomics. Samples from three sources (in this case a pre-ripe and a ripe tomato fruit and an internal loading control) are covalently labeled with one of the three Cy dyes (Cy2, Cy3 and Cy5); the samples are pooled, and then separated on a single 2-D gel. Imaging of the gels at different wavelengths, corresponding to the emission spectra of the three dyes, allows a quantitative and qualitative comparison of the protein populations in the original samples and the differential images may be readily analyzed to determine statistical differences (Rose et al., 2004).

An alternative to the 2-D gel approach is LC-MS/MS analysis of total peptide digests. Since peptide HPLC separation and ionization in the mass spectrometer is not highly reproducible, peak intensities of the same peptide from different experiments are difficult to compare. To circumvent these limitations, stable isotope-labeling techniques have been developed, based on labeling of protein extracts from two or more samples with stable isotopes. After labeling, proteins are digested, and peptides are mixed and analyzed in one and the same experiment, using the isotopic difference to determine the origin of the peptide. The same peptides from different labeled samples have identical chemical properties, so they will behave the same during separation and ionization. Therefore, the peak intensities correlate with peptide abundance and protein levels. Compared to DIGE, gel-free methods are more sensitive, and so they are more suited to analyze low abundant proteins. Two classes of labeling are distinguished: *in vivo* metabolic labeling and *in vitro* chemical labeling.

Metabolic labeling takes advantage of the biosynthetic incorporation of isotope-labeled nutrients or amino acids into proteins (Figure 3). SILAC (Ong et al., 2002; Ong and Mann, 2006) or stable isotope-labeled amino acid in cell culture seemed inapplicable to plants because of the autotrophic nature of plant cells and as a consequence the low efficiency of incorporation. Recently however, an incorporation efficiency of 80% was reached in *Arabidopsis* using ^{13}C-arginine (Gruhler et al., 2005). To achieve a more accurate *in vivo* quantitative representation of a plant proteome, two new methodologies were developed: SILIP or stable isotope labeling *in planta* allows soil-based grown plants to be efficiently labeled using a ^{14}N/^{15}N isotope coding strategy (Schaff et al., 2008), while HILEP or hydroponic isotope labeling of entire plants uses hydroponic media containing ^{15}N inorganic salts as the sole nitrogen source (Bindschedler et al., 2008). Both approaches reached near to 100% ^{15}N-labeling of proteins in different tissues of respectively tomato plants and *Arabidopsis*.

In contrast, chemical labeling is performed after protein extraction. A typical example of chemical labeling is ICAT (Adam et al., 2002) (Figure 3). The method relies on covalent modification of Cys residues with an isotope-labeled affinity tag. This tag consists out of a thiol-specific reactive group, an isotope tag composed of either ^{12}C or ^{13}C, and a biotin

affinity tag. After differential labeling of protein extracts from different samples, proteins are pooled, digested, and tagged cysteine-containing peptides are enriched from the peptide mixtures by avidin affinity chromatography, greatly decreasing the complexity of the sample (Dunkley et al., 2004; Hartman et al., 2007). One limitation of ICAT is the low frequency of Cys amino acids in proteins, leading to false negatives.

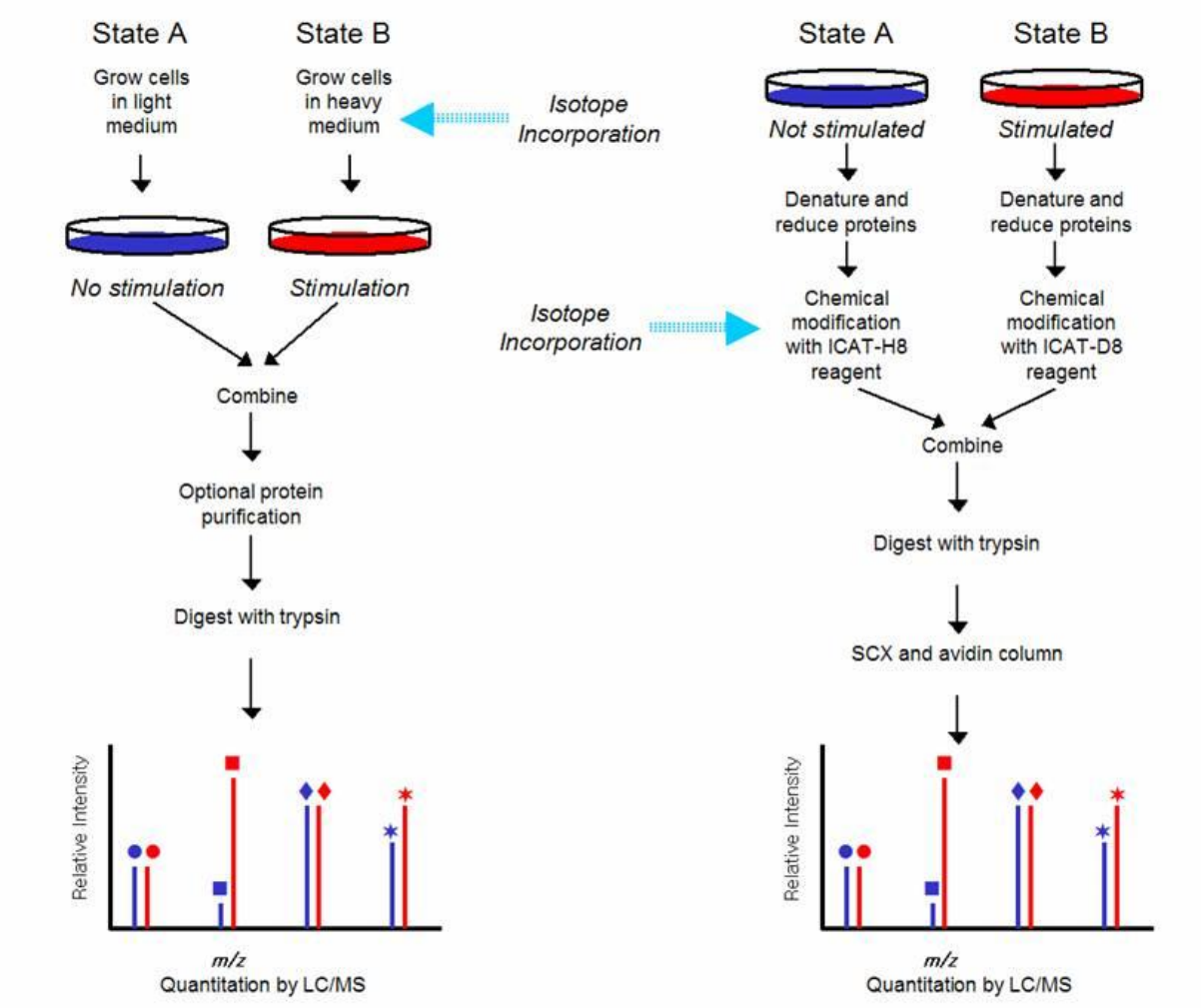


Figure 3: Illustration showing the principle of SILAC and ICAT (source: http://www.proteome.re.kr).

To overcome this problem, iTRAQ was developed (Washburn et al., 2001) (Figure 4). The principle remains the same, but now primary amines, present in every trypsin digested peptide, are targeted. The iTRAQ tag is a multiplex isobaric tag that contains an isobaric linker that, upon fragmentation of the peptide, releases a characteristic mass reporter that appears in the immonium ion region of tandem mass spectra. The major advantage of iTRAQ

is that up to four samples can be compared simultaneously in one experiment. As all peptides are labeled, high proteome coverage is reached. However, comparisons of total protein digests are more complex than those from ICAT, and low abundant peptides may be obscured from detection (Peck, 2005; Chen and Harmon, 2006).

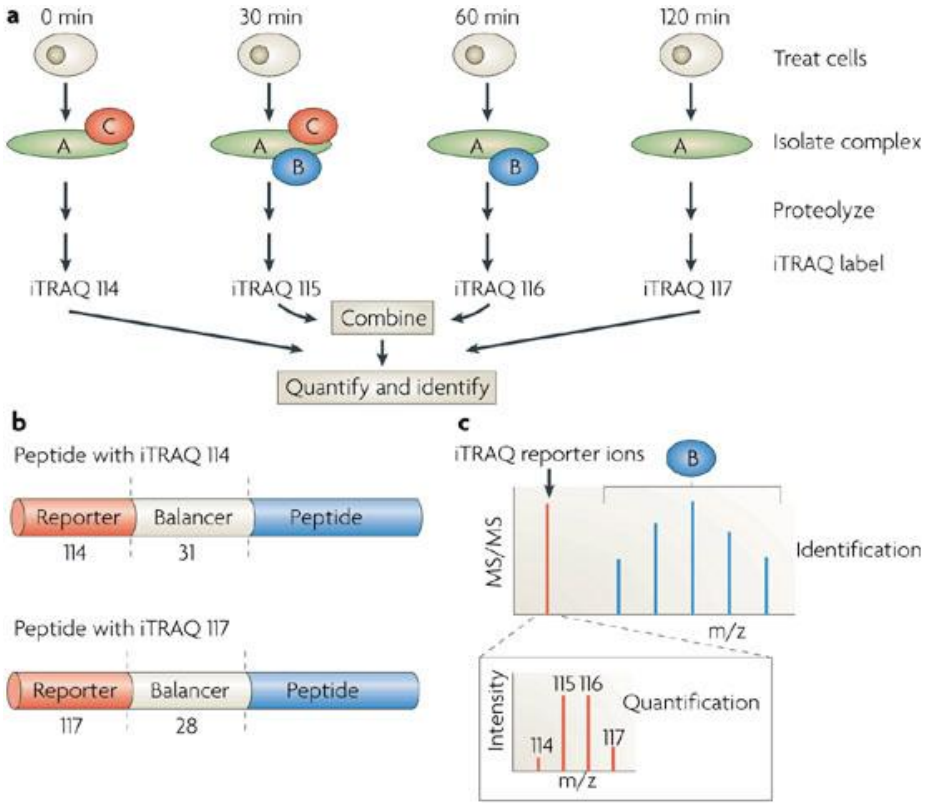


Figure 4: Principle of iTRAQ. a) Up to 4 samples can be analyzed. Isobaric tags are chemically added to the N-term. of every peptide. b) iTRAQ is isobaric, to maintain a constant mass, the reporter moiety is separated from the peptide by a balancer group. The reporter and balancer groups fragment in the collision cell of the mass spectrometer during MS/MS, and (c) the intensity of the reporter ions is monitored (Gingras et al., 2007).

The former quantitative proteomic studies compare 'relative' amounts of proteins between different samples. Recently, a technology was developed that allows 'absolute' protein quantifications and comparisons (Gerber et al., 2003). AQUA uses chemically synthesized peptides labeled with tags containing heavy isotopes as internal standards for quantification of unlabeled peptides derived from parent proteins.

Because both gel-based and LC-MS/MS based methods all have their own limitations and advantages, they are complementary and preferentially both methods are used to reach the highest coverage.

1.4 Post-translational modifications

After translation, the range of functions of a protein is extended by post-translational modifications. PTMs modulate protein activity, localization, stability and complex assembly.

Global PTM analyses deal with intrinsic difficulties. First, a broad spectrum of PTMs such as the removal of signal peptides, processing of precursor polypeptides, and modification of amino acids has been reported (Aebersold and Goodlett, 2001). The number of possible amino acid modifications exceeds 450 (Garavelli, 2004), including attachment of functional groups (by phosphorylation, glycosylation, acetylation, myristoylation, palmitoylation, methylation, sulfation, nitrosylation, prenylation,…), addition of other proteins or peptides (e.g. ubiquitination and SUMOylation), modification of the chemical nature of amino acids (e.g. citrullination), or structural changes like the formation of disulfide bridges and proteolytic cleavage. Second, many PTMs are regulatory and reversible and involve low abundance proteins, present only at specific time points during the cell cycle, or at certain stages of plant development. Furthermore, peptides carrying modifications deal with low MS ionization and fragmentation efficiencies. And finally, most PTMs cannot be predicted accurately using bioinformatic tools, so currently, we depend almost entirely on empirical data to determine and map PTMs.

Most methodologies to detect PTMs rely on specific biochemical enrichment of modified proteins and peptides, to increase the detection sensitivity by MS.

1.4.1 Phosphorylation

Among all PTMs, phosphorylation is the most widely studied and best understood. Most cellular processes are regulated by dynamic phosphorylation events on serine, threonine, or tyrosine residues. The importance of phosphorylation is reflected by the large number of protein kinases (about 1000) and phosphatases (about 300) present in *Arabidopsis* (Kersten et al., 2006; Peck, 2006). Up to 33 different cyclin dependent kinases (CDKs) with a possible role in the cell division cycle have been identified in *Arabidopsis* (Vandepoele et al., 2002;

Menges et al., 2005). Furthermore, at least 50 members of two-component systems are predicted to be present in *Arabidopsis*. These two-component systems are involved in plant signal transduction, in particular hormone signaling, red-light perception, circadian rhythms and make use of histidine kinases, response regulators, and histidine-containing phosphotransfer proteins.

For many years, the study of phosphorylated proteins relied on the detection of a mobility shift on 2-D gels. This shift is caused by a decrease of the pI upon phosphorylation. Recently, various methods have been developed to study phosphorylations more effectively. In most of them, subproteomes are analyzed, to enrich for low abundance proteins. Proteins are digested with trypsin and phosphopeptides are selectively isolated. Immobilized metal affinity chromatography (IMAC) with Fe^{3+} or Ga^{3+} ions (Ficarro et al., 2002; Nuhse et al., 2004) or a combination of strong cation exchange and IMAC (Villen and Gygi, 2008) are often used for this purpose, however improved results were recently obtained using TiO_2 microcolumns (Pinkse et al., 2004; Larsen et al., 2005). Other methods for enrichment rely on functional modification of the phosphate group with sulfhydryl groups (Zhou et al., 2001) or substitution with biotin (Oda et al., 2001). A combination of SILAC along with a two-step strategy for phosphopeptide enrichment (cation exchange and IMAC + TiO_2) and high mass accuracy mass spectrometry applied on a human cell line, allowed mapping of 14265 unique phosphorylation sites and identification of more than 1,000 proteins with increased phosphorylation in mitosis, including many known cell cycle regulators (Dephoure et al., 2008). Data related to plant phosphorylations are deposited in databases maintained by PlantP (http://plantsp.genomics.purdue.edu/) or by PhosPhAt (http://phosphat.mpimp-golm.mpg.de/).

1.4.2 Ubiquitination

Ubiquitination, a key regulatory mechanism to control protein stability, localization, and activity in plants, has gained much interest in recent years. Ubiquitin profiling also involves enrichment steps: some use GST-tagged ubiquitin binding domains to isolate ubiquitinated proteins subsequently analyzed by MudPIT (Maor et al., 2007), others directly target ubiquitinated proteins by tagging the ubiquitin peptide with a HB-tag (Tagwerker et al.,

2006). The HB-tag allows for purification under denaturing conditions preventing loss of ubiquitination due to ubiquitin hydrolase activity.

A wide variety of other plant specific PTM analyses are reviewed in Hirano (2004), Peck (2005), and Rossignol (2006).

1.5 Protein-protein interactions

Another major objective of functional proteomics is the study of protein-protein interactions. Proteins most often do not work as isolated, monomeric entities. Instead, they function through dynamic time- and space-dependent interactions with other proteins or molecules (e.g. DNA, RNA, lipids, and metabolites). Through these interactions, they form multi-protein complexes that assemble, store, and transfer biological information (Alberts, 1998). Within a complex, each protein may have a specialized function that contributes to the overall function of the complex (Bauer and Kuster, 2003). Thus, the role of a protein can be inferred from its functional context provided by associated proteins, which may have a known function. Even when studying proteins of known function, novel insights can be obtained from describing their molecular environment. This principle of functional annotation is sometimes referred to as the 'guilt by association' concept (Drewes and Bouwmeester, 2003).

1.5.1 Protein complexes

Sometimes proteins assemble into huge molecular machines. These are involved in a wide variety of biological processes, e.g. DNA replication and transcription, RNA processing (spliceosome), protein synthesis (ribosome) or destruction (26S proteasome, anaphase promoting complex). Other molecular machines form giant structural complexes (microtubuli) or orchestrate enzymatic activities in metabolons.

Next to these, less macroscopic protein complexes are ubiquitous in the cell and execute more subtle or diverse functions, often involved in signal transduction and regulatory pathways. Cyclin Dependent Kinases (CDKs) for example, are mostly detected in association with cyclins, governing substrate specificity, and with other regulatory subunits, modulating their activity, stability and localization.

Inside the cell, proteins do not collide through constant trial and error in a diffusion-dependent manner. In contrast, protein complex assembly is very well orchestrated and regulated to ensure the efficient execution of biological processes. It is a very dynamic process involving PTMs and energy-driven conformational changes assisted by chaperones (Kurucz et al., 2002). Protein complexes are the result of coordinated gene expression, concerted translation and assembly as well as transport, activity and degradation (Gavin and Superti-Furga, 2003).

Once they are assembled, protein complexes can be further regulated by many mechanisms. Internal or external signals can alter the local concentration of individual protein components, leading to association or dissociation. PTMs or binding to metabolites can provoke 3-D conformational changes altering the affinity, co-operativity and kinetic parameters of the interactions, determined by the physicochemical and geometrical interface properties (Berggard et al., 2007).

Another important feature of multi-protein complexes is their modularity (Gavin et al., 2006). Quite often, proteins participate to more than one complex. Different complexes use often the same 'core' proteins, while functional diversity is obtained by attachment of specific proteins or 'modules', groups of proteins always found together. Core proteins are usually linked by very stable permanent interactions. In contrast, attachments are often linked to the core by interactions with a low lifetime, transient interactions.

To date, most protein-protein interactions in plants were analyzed through yeast two-hybrid or other genetic approaches (See below) and only a limited number of research groups were successful in isolating plant protein complexes. The few examples of protein complexes that were purified from plant cells by tandem affinity purification are described in chapter 3. A few others were purified in a different way. For example, the Mediator complex, a central co-regulator of transcription, was isolated from *Arabidopsis* cell suspension cultures through a combination of ion chromatography and immunoprecipitation (Backstrom et al., 2007). The Cop9 signalosome was both isolated from *Arabidopsis* plants by affinity purification through either epitope tagging (Menon et al., 2005) or TAP tagging (Rubio et al., 2005). This multi-protein complex is highly conserved and plays a key role in the ubiquitin/26S

proteasome proteolytic regulatory pathway. The 26S proteasome was previously isolated from *Arabidopsis* plants through sequential anion exchange and size exclusion chromatography (Yang et al., 2004). In another study, sucrose-density ultracentrifugation from digitonin-treated mitochondrial fractions allowed purification of the respiratory-chain I + III_2 supercomplex (Dudkina et al., 2005). Another membrane complex involved in the brassinosteroid signaling pathway was immunoprecipitated using anti-GFP antibodies targeting a CFP (Cyan fluorescent protein) fused to the Somatic Embryogenesis Receptor-like Kinase1 (SERK1) from *Arabidopsis* seedlings (Karlova et al., 2006).

1.5.2 Protein-protein interaction analysis

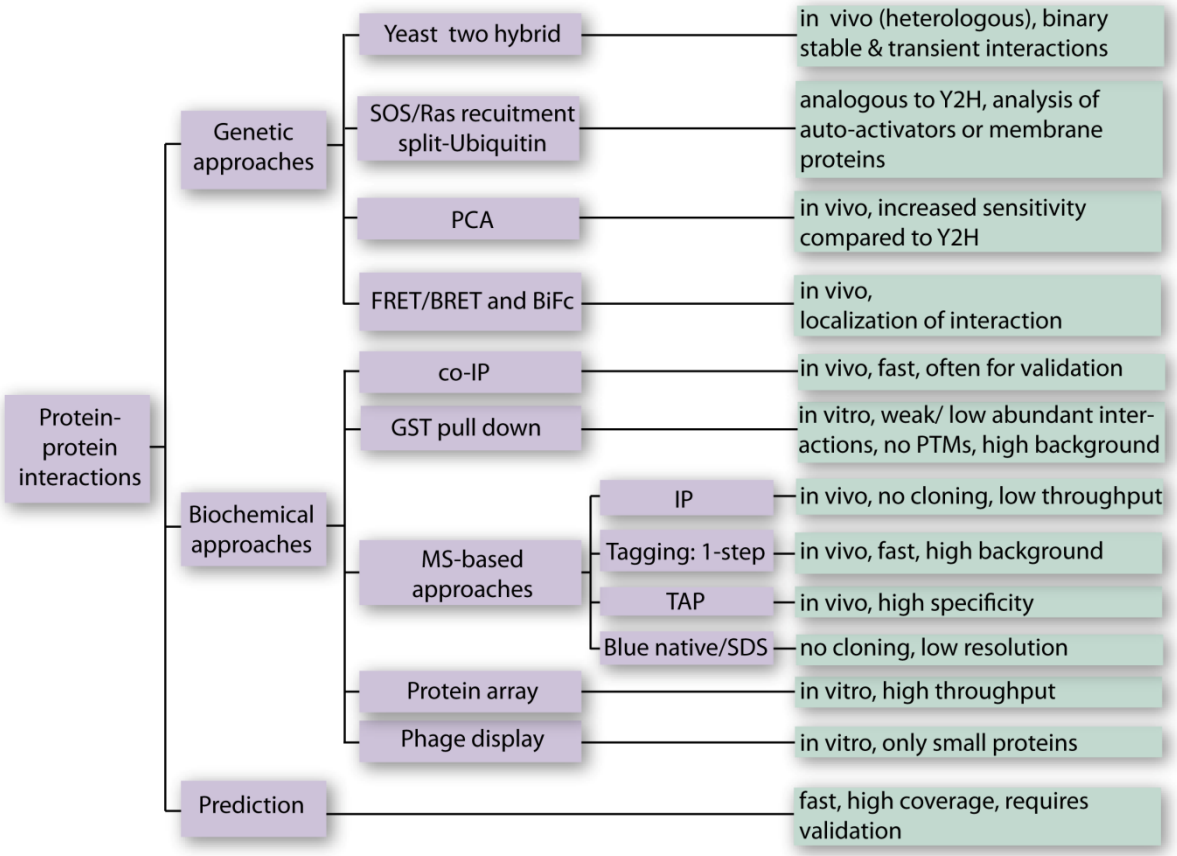


Figure 5: Overview of protein-protein interaction detection methods (purple) discussed below. Most striking advantages or limitations are shown in green boxes.

The analysis of protein complexes and protein-protein networks allows the functional annotation of gene products, and is therefore of central importance in biological research. The last two decades, a wide variety of methods were developed to study protein-protein

interactions (Figure 5). Some are based on genetic approaches, while others rely on biochemical approaches. Recently, also computational methods have proven their utility in predicting PPIs. It is important to distinguish methods that allow mapping of binary interactions, from methods that study protein complexes, as the latter can delineate both direct binary and indirect interactions. For this reason, one should be careful when comparing results obtained with different methods and when validating interactions with different approaches. In general, genetic approaches are considered as technologies that allow mapping of binary interactions, however one can not exclude that endogenous proteins assist in the interaction. Protein arrays and phage display on the other hand clearly analyze binary interactions, while the MS-based methods and co-IP interrogate both direct and indirect interactions, as these are based on complex isolation. Finally, GST pull down can study both types, depending on the applied approach. Furthermore, one has to distinguish two types of interactions, real physical interactions and functional relationships between proteins. Proteins that are functionally related are involved in the same biological pathway, but do not necessarily interact. Deduction of functional relationships is more a genetic tool and is often used as a validation tool to confirm observed interactions or in the prediction of protein interactions. Below, the main emphasis is put on the analysis of physical interactions as this is the topic of this thesis.

1.5.2.1 Mapping physical interactions

Genetic approaches to map PPIs

Yeast two-hybrid

The yeast two-hybrid (Y2H) system is a genetic *in vivo* assay in yeast that detects binary (direct) physical interactions. As such, analysis of interactions between proteins not belonging to yeast are conducted in a heterologous system. The system requires the construction of hybrid genes that encode a DNA-binding domain of the GAL4 transcription factor fused to a target protein, the 'bait', and a GAL4-derived activation domain fused to a second protein, the 'prey'. Upon interaction of bait with prey, the DNA-binding domain and the activation domain are brought into close proximity, reconstituting a functional transcription factor (TF) that stimulates the activation of a reporter gene or a selectable

marker (Fields and Song, 1989) (Figure 6). Once the TF is reconstituted, the interaction is stabilized, and therefore, it is especially well suited for the identification of weak, transient interactions.

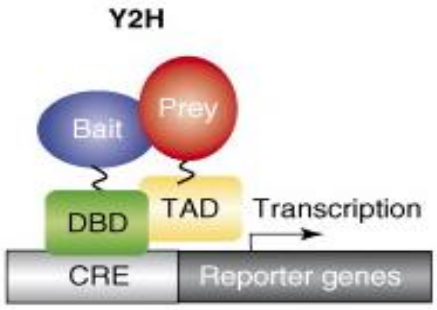


Figure 6: Yeast two-hybrid (Y2H) detects PPI by way of one or more transcriptional reporters following reconstitution of the DNA-binding domain (DBD) and Trans-activation domain (TAD) of a split transcription factor. CRE = Cis-regulatory element (Morsy et al., 2008).

Both homodimeric and heterodimeric interactions can be detected, however, cooperative interactions between more than two proteins are not detectable. Generating mutants, the screen is also often used to pinpoint amino acid residues or motifs that are critical for a specific PPI. The system is rather inexpensive compared to biochemical approaches. It is easily automated and, as a consequence, it evolved rapidly from an assay for the detection of an interaction between two known proteins, to a comprehensive genome-wide screening assay, in which a library of full-length proteins is screened against a single bait protein or even against itself. This approach gave rise to comprehensive protein-protein interaction maps in *Saccharomyces cerevisiae* (Fromont-Racine et al., 1997; Uetz et al., 2000; Ito et al., 2001; Yu et al., 2008), *Caenorhabditis elegans* (Walhout et al., 2000; Reboul et al., 2003; Li et al., 2004), *Drosophila melanogaster* (Giot et al., 2003) and human (Colland et al., 2004; Rual et al., 2005; Stelzl et al., 2005). In plants however, high-throughput Y2H screening is lagging considerably behind, even though the methods (Fang et al., 2002; Chern et al., 2007; Tardif et al., 2007) and cDNA collections are available (Gong et al., 2004; Hilson, 2006; Underwood et al., 2006). Nevertheless, a few large-scale projects have been initiated (Kersten et al., 2002; Paz-Ares and The Regia, 2002), and Y2H allowed comprehensive mapping of e.g. *Arabidopsis* MADS box transcription factors (de Folter et al., 2005).

Unfortunately, like most technologies, Y2H encounters some major drawbacks: first of all, the interactions occur in the nucleus, causing problems for membrane proteins and certain other protein classes. Second, transcription factors and other proteins often auto-activate transcription, giving rise to high false positive error rates (typically 5-10%). Third, ectopically

expressed proteins quite often do not undergo the required PTMs necessary for interaction. Finally, it only reveals binary interactions, necessitating the identification of the protein complexes *post hoc* via bioinformatics effort. Alternatively, Y2H can be used to determine the architecture of affinity-purified protein complexes.

Other two-hybrid assays

To overcome some of the limitations associated with traditional Y2H, variations around the same principles were developed (for reviews see (Vidal and Legrain, 1999; Causier and Davies, 2002)). The SOS recruitment system can be applied to study interactions involving auto-activators (Aronheim et al., 1997), while the Ras recruitment system (Broder et al., 1998), the G-protein-based screening system (Ehrhard et al., 2000) and the split-ubiquitin system (Stagljar et al., 1998; Thaminy et al., 2004) (Figure 7) are developed to detect interactions between membrane proteins. The Y2H method has also been adapted to function in bacterial and mammalian cells (Figeys, 2002).

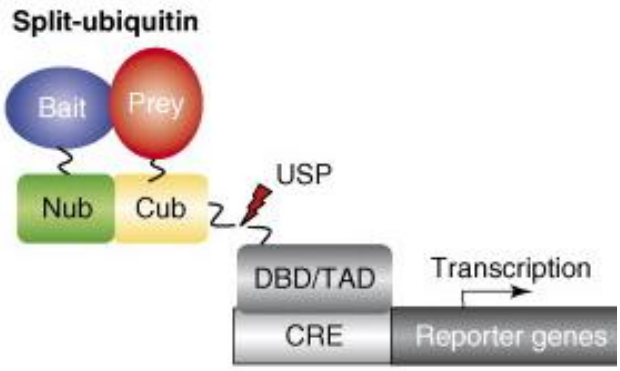


Figure 7: Split-ubiquitin detects PPI by release of a transcription factor tethered to a membrane, following activation of a ubiquitin-specific protease (USP) resulting from the reconstitution of the N-terminal half of ubiquitin (Nub) and the C-terminal half of ubiquitin (Cub). The released TF translocates to the nucleus where it activates a reporter gene. CRE = Cis-regulatory element (Morsy et al., 2008).

Other variations include protein-fragment complementation assays (PCA) (Michnick et al., 2000), that rely on the reconstitution of an enzyme (Figure 8), e.g. DHFR (Pelletier et al., 1998; Remy et al., 2007a), β-lactamase (Wehrman et al., 2002; Remy et al., 2007b), or luciferase (Fujikawa and Kato, 2007) leading to a measurable enzymatic assay.

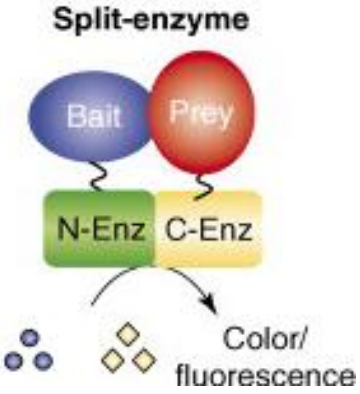


Figure 8: Split-enzyme results in the generation of color from a chromogenic substrate (Blue circles) or fluorescence from a fluorescence substrate (yellow diamonds) upon reconstitution of the enzyme (e.g. DHFR, β-lactamase or luciferase) facilitated by bait-prey interaction; N-Enz, N-terminal half of enzyme. C-Enz, C-terminal half of enzyme (Morsy et al., 2008).

These systems have important advantages compared to Y2H: in general, the PCAs can be performed in most cell types or in diverse cell compartments, the test proteins are expressed at low-level, the signal is a direct result of the interaction, and it is enzymatically amplified, leading to increased sensitivity. However, like all technologies, each has its own specific limitations. For example, a major drawback of the β-lactamase PCA assay is that the substrate is not taken up equally by different cell types and, in many cases, not at all (e.g., plant and yeast). Recently, a genome-wide DHFR PCA assay allowed mapping of the yeast protein-protein interactome (Tarassov et al., 2008).

Visualization of protein interactions in living cells

In plants, the DHFR PCA was used with a fluorescent probe allowing direct and quantitative visualization of protein interactions (Subramaniam et al., 2001). However, the most widely used approach for visualization of protein-protein interactions in living cells is fluorescence resonance energy transfer (FRET) between spectral variants of the green fluorescent protein (GFP) fused to the associating proteins (Hink et al., 2002) (Figure 9). FRET allows qualitatively and quantitatively monitoring of PPIs using advanced fluorescence microscope techniques including wide-field, confocal, multiphoton, spectral imaging, lifetime, and correlation spectroscopy (Chen and Periasamy, 2007). However, the maximum distance over FRET can take place is ca. 7 nm, provoking problems when studying proteins in large complexes. Furthermore, FRET obligatorily necessitates fluorescence excitation with its concomitant problems of photobleaching, autofluorescence, phototoxicity, and undesirable stimulation of photobiological processes. A sister technique, bioluminescence resonance energy transfer

(BRET), avoids these problems because it uses enzyme-catalyzed luminescence (Subramanian et al., 2004; Xu et al., 2007).

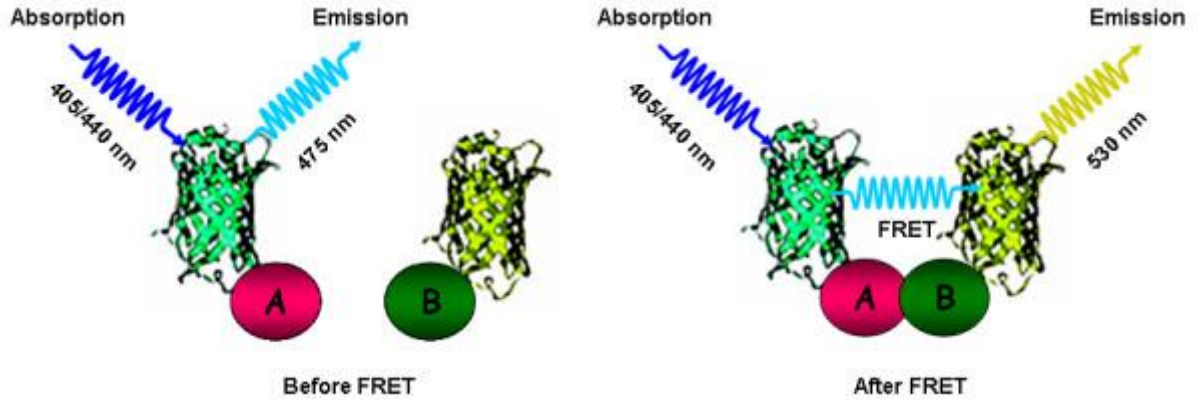


Figure 9: FRET between cyan fluorescent protein (CFP) as a donor fused to protein A and yellow fluorescent protein (YFP) fused as an acceptor to protein B. Under favorable spatial and angular conditions, interaction between A and B causes a decrease in the intensity of donor (CFP) fluorescence concomitant with an increase in acceptor (YFP) fluorescence. CFP and YFP are depicted as cyan and yellow ribbon models fused to putative interacting proteins A and B, respectively (Bhat et al., 2006).

As an alternative to FRET, bimolecular fluorescence complementation (BiFC) was developed (Hu et al., 2002). In BiFC, non-fluorescent fragments of the enhanced yellow fluorescent protein (YFP) are fused to two proteins, and when the two proteins interact, fluorescence is restored, given that the two fragments can fold properly (Figure 10). Like in Y2H, BiFC allows the detection of both stable and transient interactions: initially, the two proteins fused to the YFP fragments interact reversibly, but as soon as the YFP fragments associate, the initial complex is stabilized.

By introducing a large number of different GFP variants, the technique was extended to multicolour BiFC, allowing the direct visualization of multiple PPIs within the same cell. Next to the use of BiFC as a PPI screening assay, it also provides evidence of the intracellular locations where the protein association occurs. Several groups have used BiFC in plants (Bracha-Drori et al., 2004; Citovsky et al., 2006; Ohad et al., 2007; Citovsky et al., 2008), and complementary sets of expression vectors were generated for PPI studies in transiently or stably transformed plant cells (Walter et al., 2004). Recently, FRET was combined with BiFC to allow visualization of ternary complexes in living cells (Shyu et al., 2008).

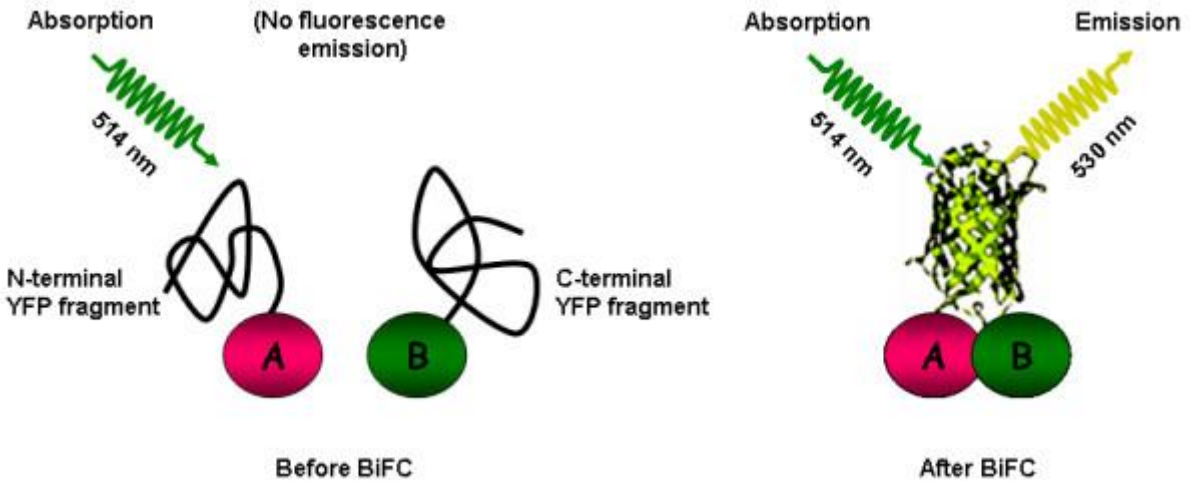


Figure 10: The scheme depicts the principle of the BiFC assay, exemplified by a split YFP fluorophore. Proteins A and B are fused to N- and C-terminal fragments of YFP, respectively. In the absence of an interaction between A and B, the fluorophore halves remain non-functional. Following interaction between A and B, a functional fluorophore is reconstituted which exhibits emission of fluorescence upon excitation with an appropriate wavelength (Bhat et al., 2006).

All methods based on visualization of PPIs, have the advantage that they screen for PPIs *in vivo*, ensuring that no PTM-dependent PPIs are missed.

Biochemical approaches to map PPIs

GST pull-down

Traditionally, PPIs were analyzed *in vitro* using immobilized recombinant proteins. The most commonly used system was based on the affinity of glutathione for glutathione S-transferase (GST). First, the protein of interest is fused to GST using *in vitro* translation methods or *E. coli* as a recombinant protein production system. Next, the recombinant fusion protein is immobilized on sepharose beads with covalently attached glutathione (Smith and Johnson, 1988). Pair wise interactions may be tested by applying a second labeled recombinant protein. Alternatively, protein extracts can be incubated with the immobilized bait protein. After washing, the precipitated complex is eluted by competition with excess of free glutathione or by boiling in the presence of SDS, and subjected to SDS-PAGE. The interaction then may be detected based on the label (e.g. radioactivity, fluorography) or in case of protein extracts via immunoblotting or MS (see below). The major advantage of this approach is that it is capable of retrieving weakly interacting or low

abundant proteins owing to the fact that large amounts of recombinant proteins are present on the column. Of course, the system has some important drawbacks: some proteins are not easily expressed in *E. coli*, PTM-dependent interactions may be missed, GST-tagging may provoke sterical hindrance, and finally, the recombinant fusion protein has to compete with the corresponding endogenous component. Nevertheless, it is an easy to use and robust approach (Bauer and Kuster, 2003). For example, *in vitro* pull-down assays in tobacco revealed that cell cycle-regulated CycD3s bind to CDKA, CDKB and to the CDK inhibitor NtKIS1a (Kawamura et al., 2006).

Co-immunoprecipitation

Co-immunoprecipitation (co-IP) was deducted from the GST pull-down principle, and, because of the high speed it provides, it is now often used to validate PPIs detected with another method. For co-IP, several strategies can be followed: first, one can use highly specific antibodies to perform the co-IP from cell lines expressing their endogenous proteins. Second, cells expressing a tagged bait protein can be used. A wide variety of tags is available, including GST, His, Flag, Myc, calmodulin-binding peptide (CBP), hemagglutinin (HA), StrepII, StrepIII, streptavidin binding peptide, … (Terpe, 2003; Arnau et al., 2006). An antibody directed against the tag can then be used during the co-IP. Finally, one can conduct co-IP experiments using cells expressing tagged versions of two putative interaction partners. Co-IP experiments usually generate significant background and therefore it is important to perform negative controls in parallel (Berggard et al., 2007).

MS-based methods

The emergence of powerful, sensitive high-throughput MS techniques, allowing the detection of peptides in the lower femtomolar range, together with the availability of comprehensive protein sequence repertoires, allowing the identification of proteins, has fuelled the development of methods employing the biochemical purification of protein complexes. MS-based approaches generally comprise two major steps: first, the isolation, fractionation, and purification of proteins, followed by the identification of proteins by MS-analyses (Chang, 2006). Protein complexes can be isolated by combining ultracentrifugation, sucrose density-gradient centrifugation, gel filtration, or ion-exchange chromatography,

however these methods are characterized by a high false positive error rate and they are not generic and must be adapted to a particular complex of interest (Chang et al., 2005; Gingras et al., 2007). The most widely used approaches for biochemical isolation of protein complexes are based on affinity purification. Usually the starting point is an undefined heterogeneous group of molecules in solution, such as a cell lysate. The protein of interest will have a well known and defined property which can be exploited during the affinity purification process. The process itself can be thought of as an entrapment, with the target protein complexes becoming trapped on a solid or stationary phase. The other molecules in solution will not become trapped as they do not possess this property. The solid medium can then be removed from the mixture, washed and the target protein complexes released from the entrapment in a process known as elution. After the protein complex is isolated, it can be subjected to either gel-based or gel-free separation followed by an MS analysis (Figure 11). An advantage of gel-based methods is that it allows estimation of the stoichiometry of the various proteins present in the purified fraction, while this information is lost with gel-free approaches. Both methods face limitations at the level of tryptic digestion: limitations inherent to gel processing are the loss of protein sample and the possible contamination with e.g. keratin. On the other hand, in-solution digestion has to deal with low digestion efficiencies for proteins present in low concentration. To overcome this limitation, the 'Proteome Reactor' was developed. This is a micro-fluidic processing device that enables efficient enzymatic digestion of affinity-purified proteins for LC-MS/MS analysis (Vasilescu et al., 2007). In-gel digested proteins are usually spotted on a MALDI-plate and analyzed via MALDI-TOF or MALDI-TOF/TOF, and proteins are identified respectively through peptide mass fingerprinting or through database searching with the sequenced peptides (Feng et al., 2008). Peptides from in-solution digested proteins are mostly analyzed via MudPIT: after 2D-chromatographic separation of the peptides, they are ionized via ESI and analyzed via MS/MS, followed by database searching for protein identification (Feng et al., 2008).

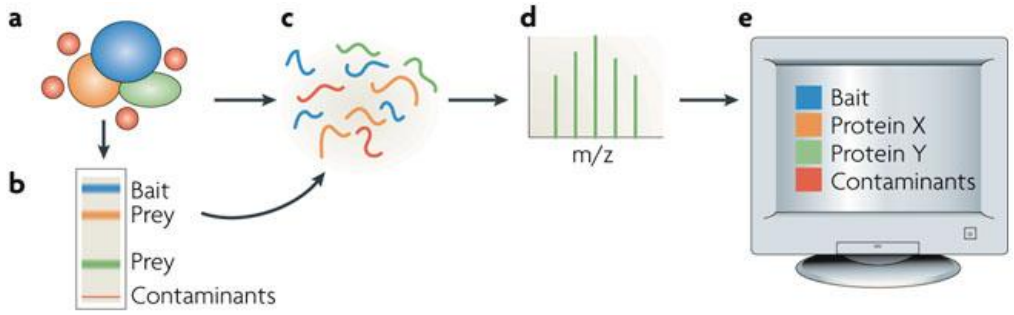


Figure 11: Analysis of protein complexes using mass spectrometry. (a) First, the protein complex of interest is isolated. Proteins are separated on gel (b), digested and analyzed by MS (d). Alternatively, with the gel-free approach (c), the complex protein mixture is directly digested and peptides are separated by liquid chromatography, followed by MS analysis. Finally, database searches allow protein identification (e) (Gingras et al., 2007).

The classical affinity chromatography-based approaches to purify protein complexes use recombinant proteins (GST pull-down) or specific antibodies against a protein of interest (immunoprecipitation). More generic approaches emerged relying on the fusion of the target protein to a certain affinity handle.

Immunoprecipitation

In a typical immunoprecipitation (IP) experiment, a protein complex is purified from a cell lysate by affinity chromatography with an immobilized antibody that specifically recognizes an epitope of one known component of the complex (Figure 12a). Next, non-specifically bound proteins are removed by extensively washing, and subsequently, the complex is eluted from the resin prior to protein identification by MS. As there is no need to ectopically express proteins, this approach resembles without doubt the closest to the physiological conditions of the cell. Furthermore, if a specific antibody of good quality is available, this approach is probably the fastest one, as no cloning is required. But, as an individual antibody is needed for every bait protein, and as purification conditions generally need to be optimized for any given protein complex, IP is not readily applied in a high-throughput mode, although antibody-fragment producing platforms could be assessed for this purpose (Eeckhout et al., 2004; Uhlen et al., 2005). Furthermore, most antibodies, even monoclonal ones, exert a certain degree of cross-reactivity, and consequently unrelated proteins and

complexes can be co-purified. Also very abundant proteins might stick to the resin or the precipitated complexes. To reduce the number of false positives, high stringency wash steps can be included prior to elution, but this will inevitably lead to dissociation of weakly bound components. Another technical problem associated with this approach is antibody bleeding from the resin during elution. If these antibodies are present in high amounts in the eluate, they can cause problems later during MS analysis by masking other proteins. Antibody bleeding can be partly prevented by cross-linking the antibody covalently to the resin or by soft, but often inefficient, peptide elution. IP can also be used for affinity depletion of very abundant proteins, e.g. rubisco in plants or albumin in mammalian serum, to study proteins at the lower end of the dynamic range (Peck, 2005).

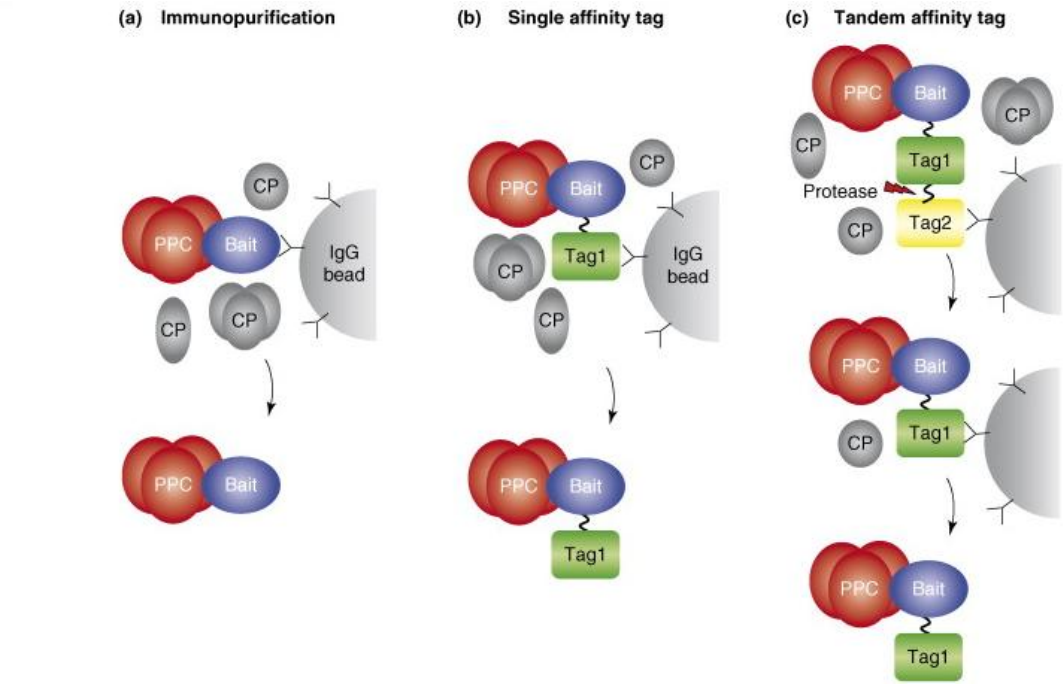


Figure 12: Methods for isolation of protein complexes by affinity purification. (a) Immunoprecipitation with immobilized antibody against the protein of interest (bait). (b) One-step purification through isolation with immobilized antibody against a universal protein tag. (c) Two-step purification using two successive affinity steps. CP = contaminating proteins, PPC = prey protein complex (Morsy et al., 2008).

Protein tagging

To overcome problems associated with the former two approaches, more elegant and/or generic purification methods were developed based on the expression of a tagged bait

protein. Proteins targeted for purification with their associated partners are modified by addition of a tag suitable for affinity purification (Figure 12b). Fusion of the target sequence to the tag is accomplished using standard DNA cloning techniques. Subsequently, the recombinant gene is introduced and expressed in the host organism via transformation with a plasmid containing the tagged gene or more elegantly by gene replacement mediated by site-specific *in vivo* recombination. To date, many different affinity purification tags are available (Table 1) (Hearn and Acosta, 2001; Terpe, 2003; Arnau et al., 2006), however for many of them the corresponding resins generate high background levels and/or exert low affinity for their corresponding ligands, and as a consequence, low and medium abundance proteins are recovered in low yield.

Table 1: List of commonly used affinity tags, together with their size in amino acids (AA) and some purification-related comments.

Tag	Size (AA)	Comments
His-tag	5–15	Purification under native or denaturing conditions
FLAG	8	Calcium-dependent, mAb-based purification
Strep-tag II	8	Modified streptavidin (Strep-Tactin), elution with biotin analog
Strep-tag III	28	Double repeat of Strep-tag II with improved binding
SBP	38	Binds to streptavidin (K_d of 2,5 nM), elution with biotin analog
HA-tag	9	Influenza virus hemagglutinin tag, Ab-based purification
Softag1, Softag 3	13, 8	Recognized by polyol-responsive mAb
c-myc	10	mAb-based purification
T7-tag	11–16	mAb-based purification
S-tag	15	S-protein resin affinity purification, derived from RNase A, harsh elution
Elastin-like peptides	18–320	Protein aggregation by temperature shift, intein used to remove tag
Chitin-binding domain	52	Binds only insoluble chitin, elution with DTT, β-MercaptoEtOH or cysteine
Thioredoxin	109	Affinity purification with phenylarsin oxide–Sepharose, improved solubility
Xylanase 10A	163	Cellulose based capture, elution with glucose
GST	201	Glutathione or GST-Ab affinity
MBP	396	Amylose affinity purification, improved folding and solubility
CBP	26	Calcium-dependent binding to calmodulin, elution with EGTA or EDTA

The recovery of interacting proteins is a function of their binding constant and abundance, their solubility and concentration in the cell extract, and their stability meaning both intrinsic stability of the interacting proteins under experimentally conditions, and their resistance to attack by enzymes in the extract that would destroy them or disrupt their associations. Epitope tags like the hemagglutinin (HA) tag, the Flag tag or the c-myc tag are derived from linear epitopes possessing high affinity for a specific antibody, allowing purification with immobilized antibodies. In general, these epitope tags are small and allow mild elution with an excess of free peptide to displace the bait protein and bait-associated proteins, but not proteins adsorbed non-specifically to the antibody or immobilization matrix, introducing a high degree of specificity to this approach. However, these competitive elution steps are often inefficient as was reported for the HA and c-myc tag when applying plant extracts (Earley et al., 2006), and high levels of free peptide are incompatible with LC-MS. An alternative approach for specific and mild elution of captured protein complexes is the use of site-specific proteases to cut at sites engineered adjacent to the tag, although the purification protocol becomes longer and the sample gets contaminated with the protease. The robustness of the Flag tag is further demonstrated by the first systematic identification of protein complexes in yeast (Ho et al., 2002) and in human cells (Ewing et al., 2007). A major drawback of epitope tags is that the corresponding immuno-affinity resins are often expensive. Cheaper solutions are provided with e.g. the polyhistidine (His_6) tag, a small peptide tag allowing purification based on the affinity of Histidine for Nickel ions (IMAC). A large-scale comprehensive pull-down assay was performed using a His-tagged *Escherichia coli* ORF clone library (Arifuzzaman et al., 2006). Other cheap purification systems rely on the interaction between biotin and Streptavidin, one of the strongest non-covalent interactions observed in nature with a dissociation constant of ± 10^{-15} M. The Strep-tag II is a small Streptavidin-binding peptide (Witte et al., 2004). However, its affinity for Streptavidin is rather weak (13 µM) (Keefe et al., 2001). Stronger binding can be achieved with Strep-Tactin, a mutated version of Streptavidin (Skerra and Schmidt, 2000), with the Strep-tag III (Junttila et al., 2005), i.e. a tandem repeat of the Strep-tag II separated by a Glycine/Serine spacer, or the SBP-tag, a 38 amino acid long peptide with affinity in the lower nanomolar range (Keefe et al., 2001; Wilson et al., 2001), while soft elution is still possible with (desthio)biotin.

Finally, protein or protein domain fusion tags can be used, e.g. the maltose-binding protein (MBP) or the previously described GST. These larger tags often increase folding and solubility, however, as they often interfere with correct protein complex assembly due to their bigger size, they are less suitable for protein interaction analysis.

Tandem affinity purification

The TAP tag method

As one-step purification often has to deal with high background (low purification) levels, a strategy for the isolation of native protein complexes was developed. Since the procedure implies two consecutive affinity purification steps, it was named tandem affinity purification (Rigaut et al., 1999) (Figure 12c). The technology was originally developed for *Saccharomyces cerevisiae*. The use of two independent affinity steps greatly enhances the specificity of the purification procedure. Preferentially, the fusion protein is expressed at close to natural levels to minimize assembly of non-physiological complexes, and consequently, a combination of high-affinity tags is required. In yeast, homologous recombination allows fusion of the TAP tag to the endogenous gene, so the original expression levels are maintained here, and when using haploid cells, there will be no competition of the fusion protein for complex assembly with an endogenous counterpart. Originally, several affinity tags were tested but the best results were obtained with two IgG-binding units of protein A of *Staphylococcus aureus* (ProtA) and the calmodulin binding peptide (CBP) (Figure 13). While the CBP tag allows efficient elution under close to physiological conditions, ProtA release from the IgG matrix requires denaturing conditions at low pH. This problem is solved by the addition of the specific Tobacco Etch Virus (TEV) cleavage site allowing proteolytic release under mild conditions, keeping the eluted complexes intact. The resulting TAP cassette is thus a combination of the CBP tag and the ProtA tag, separated by the TEV cleavage site, with the ProtA tag at the extreme terminus (Figure 13). Purification buffers were optimized for highest yield, while generally maintaining protein complex integrity in an environment not too highly divergent from the intracellular conditions. The TAP method involves the fusion of the TAP tag to the target protein, and the transformation of the construct into the host cell or organism.

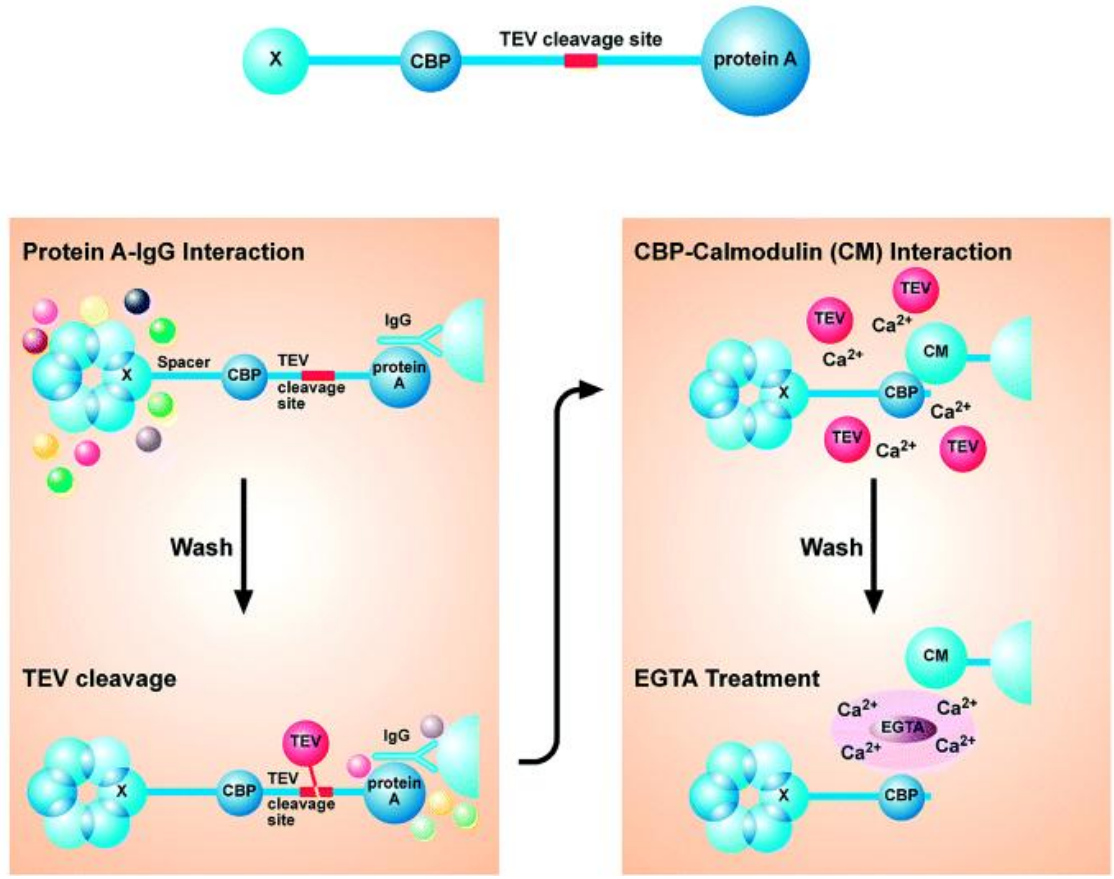


Figure 13: Diagram illustrating the originally described tandem affinity purification method. During the first purification step, complexes are isolated on a IgG resin, followed by a wash step and gentle elution with TEV protease. Subsequently, the released complexes are further purified on a calmodulin (CM) resin and complexes are finally released by addition of EGTA. X represents the bait protein of interest (Bauer and Kuster, 2003).

The fusion protein and associated components are recovered from cell extracts by selection on an IgG matrix. After washing, the TEV protease is added to elute the bound complexes. Next, the eluate is incubated with calmodulin-coated beads in the presence of calcium ions. This second affinity step is required to remove the TEV protease as well as contaminating proteins remaining after the first affinity selection. After washing, the bound material is released with EGTA, through chelation of the calcium ions and disruption of the CBP/calmodulin interaction. When C-terminal fusions fail, the N-terminal version of the TAP tag can be used in which the modules are reversed, with the ProtA tag at the extreme terminus as it will be cleaved off during the first purification step. The split-tag strategy in

which the ProtA tag and TEV site are fused to one protein and the CBP tag to a second one, is particularly useful in cases where some subunits belong to several different complexes, while the subtraction strategy can be followed when undesired complexes have to be selectively removed, by tagging one component of the undesired complex with the ProtA tag but without the TEV cleavage site (Puig et al., 2001).

Large-scale/high-throughput interaction studies

The technology was successfully transferred to a broad range of other host systems e.g. *Escherichia coli* (Butland et al., 2005), *Schizosaccharomyces pombe* (Gould et al., 2004), Trypanosomes (Estevez et al., 2001), *Helicobacter* (Stingl et al., 2008), human cell lines (Bouwmeester et al., 2004; Brajenovic et al., 2004), *Drosophila melanogaster* (Veraksa et al., 2005), *Arabidopsis thaliana* (Rohila et al., 2004; Van Leene et al., 2007), *Oryza sativa* (Rohila et al., 2006), and many others. Mapping of protein interactions on a genome-wide scale requires high-throughput methods of cloning, transformation, purification and identification, and is not yet feasible in all systems. For yeast however, these high-throughput methods are available, and consequently the first two comprehensive analysis of protein complexes were conducted in *S. cerevisiae*. One of them was based on overexpression of 493 Flag epitope tagged baits (See above) unraveling 3617 interactions among 1578 proteins. However due to the overexpression approach and the low purification, the dataset may be prone to false positives (Ho et al., 2002). Simultaneously, a second comprehensive yeast interactome was mapped by TAP of 1739 proteins revealing 589 protein assemblies (Gavin et al., 2002). This was a real landmark survey of protein interactions as it was the first screen analyzing protein interactions under close to physiological conditions on such a large scale, while previous studies used heterologous approaches as Y2H or *in vitro* techniques such as protein chips. More recently, the first genome-wide screen for protein complexes using TAP combined with MALDI-TOF was reported (Gavin et al., 2006). In this screen, 6466 yeast open-reading frames (ORFs) were tagged, allowing purification of 1993 proteins and identification of 491 distinct protein complexes. In this work, the authors suggested that some core proteins are associated with different protein modules in different complexes, enabling diversification of potential functions. In a related study, 4562 genes were analyzed by TAP leading to identification of 547 complexes, averaging 4.9 proteins/complex (Krogan et al., 2006). In this

study two methods of MS were utilized for protein identification (MALDI-TOF and LC-MSMS), leading to increased coverage and accuracy. Machine learning was used to integrate the mass spectrometry scores and to assign confidence to the protein–protein interactions. High-confidence interactions were extracted from these 2 studies and modeled into a reliable binary yeast interactome of 9074 interactions (Collins et al., 2007).

Also for *E. coli*, two comprehensive studies were conducted. One of them revealed a protein network of 5254 interactions through a combination of TAP and sequential peptide affinity tagging (See Table 2) (Butland et al., 2005), while the other one used His-tagging to map protein interactions around 4339 test baits. Comprehensive studies in higher eukaryotes as mammalia and plants are running behind because the high throughput methods can not be applied in these organisms to the same extend as in prokaryotes, although efforts are done, e.g. in humans, 338 Flag-tagged baits unraveled 6463 interactions among 2235 distinct proteins (Ewing et al., 2007).

In parallel with TAP, comprehensive protein interaction networks were obtained with genetic approaches. Y2H allowed extensive mapping of binary protein interaction networks for yeast (Uetz et al., 2000; Ito et al., 2001; Yu et al., 2008) and higher eukaryotes as *Caenorhabditis elegans* (Walhout et al., 2000; Reboul et al., 2003; Li et al., 2004), *Drosophila melanogaster* (Giot et al., 2003) and human (Colland et al., 2004; Rual et al., 2005; Stelzl et al., 2005). The authors of the most comprehensive binary map of yeast obtained so far, stated that, despite the large size of the dataset, only 20 % of the yeast binary interactome, which is estimated to consist of approximately 18000 ±4500 interactions, is mapped to date (Yu et al., 2008). Other genetic approaches like LUMIER, based on detection of luciferase tagged preys after immunoprecipitation of Flag tagged baits (Barrios-Rodiles et al., 2005), and the DHFR protein complex assay (Tarassov et al., 2008), allowed the construction of genome-wide binary interaction maps for mammalia and yeast respectively. However, overlap of binary data with data from affinity-purified complexes is low (Yu et al., 2008), which makes sense for different reasons. First of all, it is difficult to compare pairwise interactions from genetic approaches with multiprotein complex data obtained with affinity purification approaches. If two proteins are found to co-purify with affinity purification, it

does not necessarily mean that they interact in a direct, binary fashion. In addition, both approaches are prone to false negatives: transient interactions are often lost with affinity purification, while these are stabilized with Y2H or PCA. On the other hand, protein interactions dependent on specific PTMs are often missed with Y2H. So to conclude, as co-complex data and genetic approaches as Y2H interrogate a different subspace within the whole interactome, both approaches are essential to get a complete picture of cellular protein-protein interaction networks, and they should be used in parallel.

Improvement of the TAP tag method

As homologous recombination is not feasible in most organisms as it is in yeast and *E. coli*, there will be competition for complex assembly between the endogenous protein and the fusion protein, decreasing the success rate of TAP in these organisms. To make more partner subunits available for association with the target protein, depletion of endogenous protein by RNAi-mediated gene inactivation (Forler et al., 2003) was implemented. A similar strategy resides on the transformation of the transgene encoding the fusion protein in a mutant background where the endogenous counterpart is eliminated, e.g. by T-DNA insertion in *Arabidopsis* (Rubio et al., 2005). The latter approach immediately allows the determination of the functionality of the fusion protein depending on complementation of the mutant phenotype. Competition can also be increased by overexpression of the fusion protein, although this may lead to lower complex yield or the formation of non-physiological complexes e.g. with heat shock proteins or with the proteasome. In general, the traditional TAP tag can suffer from low yield and high contamination when extended to more complex organisms. This low yield is especially problematic when it is difficult or time-consuming to obtain a lot of input material, e.g. when low abundant complexes (Drakas et al., 2005) or slow growing plants are studied. Major loss during the first step is provoked by inefficient TEV cleavage, while binding of endogenous calmodulin to CBP-tagged proteins and binding of endogenous CBP to calmodulin beads accounts for loss in the second step. Moreover, the original TAP procedure often generates high background levels mainly due to contamination with calmodulin-binding proteins (Van Leene et al., 2008). That limitations are associated with the original TAP tag is further demonstrated by the wide variety of different TAP tags that were developed for higher eukaryotes in the last decade (Table 2). In mammalian cells,

improved results were obtained with the GS tag, which combines the SBP tag and the Protein G tag. A 10-fold increase in final yield with this GS tag was reported (Burckstummer et al., 2006). Evaluation of this GS tag in *Caenorhabditis elegans* (Kyriakakis et al., 2008) and in *Arabidopsis thaliana* (Van Leene et al., 2008) confirmed the GS tag's superiority. In addition, these studies revealed that it generates less background and improves expression. A detailed overview of tandem affinity purification applications in plants is given in Chapter 3 (TAP in plants: an overview).

Table 2: Overview of different multiple-affinity tags developed so far. Main differences compared to the traditional TAP tag are listed.

Modules	Characteristics	References
CBP-His_6-3xHA (CHH)	Allows 2 or 3 consecutive steps for extreme purity	(Honey et al., 2001)
S tag-TEV-ZZ	No Ca^{2+}/EGTA required, however requires harsh elution step from S protein resin	(Cheeseman et al., 2001)
6 ≠ combinations of His_6 and Flag tag	Allows optimal positioning of tags	(Kimura et al., 2003)
ProtA-TEV_2–Flag	More efficient TEV cleavage, less background	(Knuesel et al., 2003)
His_9-$3C_2$-9xMyc	Performs as good as the original TAP tag. Combined with MudPIT	(Graumann et al., 2004)
CBP-TEV-ZZZZ	No increased yield observed	(Gould et al., 2004)
CBP-TEV-3xFlag	For isolation from bacterial and mammalian cells	(Zeghouf et al., 2004)
TAPi	Original TAP tag improved for plant cells	(Rohila et al., 2004)
9xMyc-His_6-3C-ZZ (TAPa)	Analysis of metal-dependent complexes in plants	(Rubio et al., 2005)
GFP-TEV-S tag	LAP tag for localization analysis and affinity purification	(Cheeseman and Desai, 2005)
His_6-$3C_2$–GFP	LAP tag for localization analysis and affinity purification	(Cheeseman and Desai, 2005)
BT-bait-TAP	Improved yield for mammalian cells	(Drakas et al., 2005)
SBP-TEV_2-ProtG	Improved yield and expression, lower background	(Burckstummer et al., 2006)
His_6-BT-His_6 (HB)	Purification under denaturing conditions after *in vivo* cross-linking to study weak interactions	(Tagwerker et al., 2006)
2xFlag-TEV-ZZ	Improved yield for mammalian cells	(Tsai and Carstens, 2006)
StrepII-Flag	Small to reduce complex assembly interference	(Gloeckner et al., 2007)

To summarize, the TAP technology allows efficient protein complex isolation and new versions improved the success rate of TAP in other, more complex organisms. Nevertheless, weak interactions are often lost during the purification. For this purpose, the HB tag was developed, making protein complex isolation compatible with *in vivo* cross linking, to stabilize transient interactions (See Chapter 3 supplement).

Two dimensional blue native/SDS-PAGE

During Blue Native PAGE (BN-PAGE) native protein complexes are resolved by molecular mass under non-denaturing conditions (Eubel et al., 2005; Wittig et al., 2006). First, proteins are solubilized using mild non-ionic detergents such as digitonin or Triton X-100, instead of sodium dodecyl sulfate (SDS), the strong ionic detergent used to solubilize and denature proteins for separation by SDS-PAGE. After isolation, samples are incubated with the anionic dye CBB G250 immediately before analysis. The dye binds to the surface of all proteins, primarily to Arginine residues. This binding of a large number of negatively charged dye molecules to proteins facilitates migration and separation of multi-protein complexes according to molecular weight in a first dimension, and the tendency for protein aggregation is considerably reduced. While one-dimensional (1D) BN-PAGE can be used directly to study protein interactions, the method is much more powerful when combined with another separation step in a two-dimensional (2D) format (Figure 14). Proteins migrating together in the first dimension, are then further separated in a second dimension through denaturing electrophoresis or isoelectric focusing. The main advantages of using BN-PAGE for the study of protein complexes are that it is inexpensive, there is no need for cloning, no specialized equipment is necessary, and the method is compatible with most methods of downstream analysis (e.g. immunoblotting, MS). The main disadvantages are those typical of any type of native PAGE, i.e. lack of fine resolution, protein smearing because of salts, etc. in the isolation buffers, and complex dissociation during separation (Miernyk and Thelen, 2008). The method is especially useful for analysis of membrane-associated protein complexes and allowed investigation of e.g. *Arabidopsis* mitochondrial respiratory-chain complexes (Dudkina et al., 2005) and plasma membrane complexes (Kjell et al., 2004).

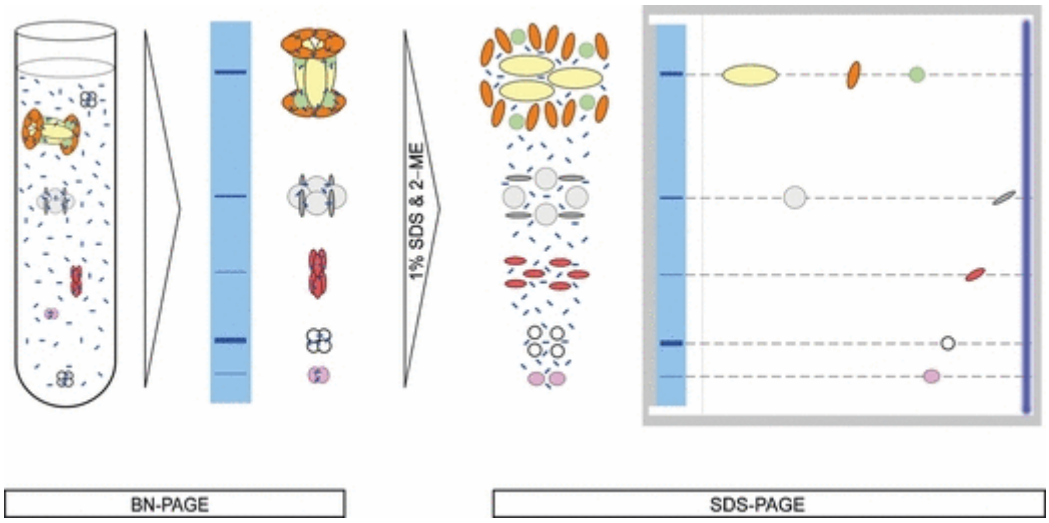


Figure 14: Two-dimensional separation of protein complexes using Blue Native PAGE in the first direction and SDS-PAGE in the second dimension (Miernyk and Thelen, 2008).

Protein arrays

In 1989, the concept of microarrays for high-throughput assays (Ekins, 1989) was introduced and soon, the first DNA microarrays were produced (Schena et al., 1995), allowing quantitative monitoring of gene expression on the transcript level. Protein arrays emerged more recently and were first used as immunoassays, utilizing antibodies as a detection method. Today, protein arrays are becoming a common tool in functional proteomics studies. They are made by immobilizing a whole library of proteins in an ordered arrangement on a high-density chip (Figure 15). This requires the availability of open-reading-frame (ORF) clone libraries, and high-throughput protein-production and purification methods. Once produced, preferentially in the host organism, and purified, the proteins are transferred to the chip using a micro-contact printer or a piezoelectric spray arrayer. In the next step, the chip is assayed by incubating them with a labeled probe. Finally, proteins interacting with the probe are detected, typically by fluorescence scanning (Kung and Snyder, 2006). So, if a protein is used as a probe, protein arrays can be used to study PPIs. In plants, several protein arrays proved their utility: an *Arabidopsis thaliana* protein array of 1690 proteins was used to identify substrates of two mitogen-activated protein kinases (Feilner et al., 2005). Calmodulin binding proteins were identified using a protein array of 1133 proteins (Popescu et al., 2007). Scanning the array with active kinases

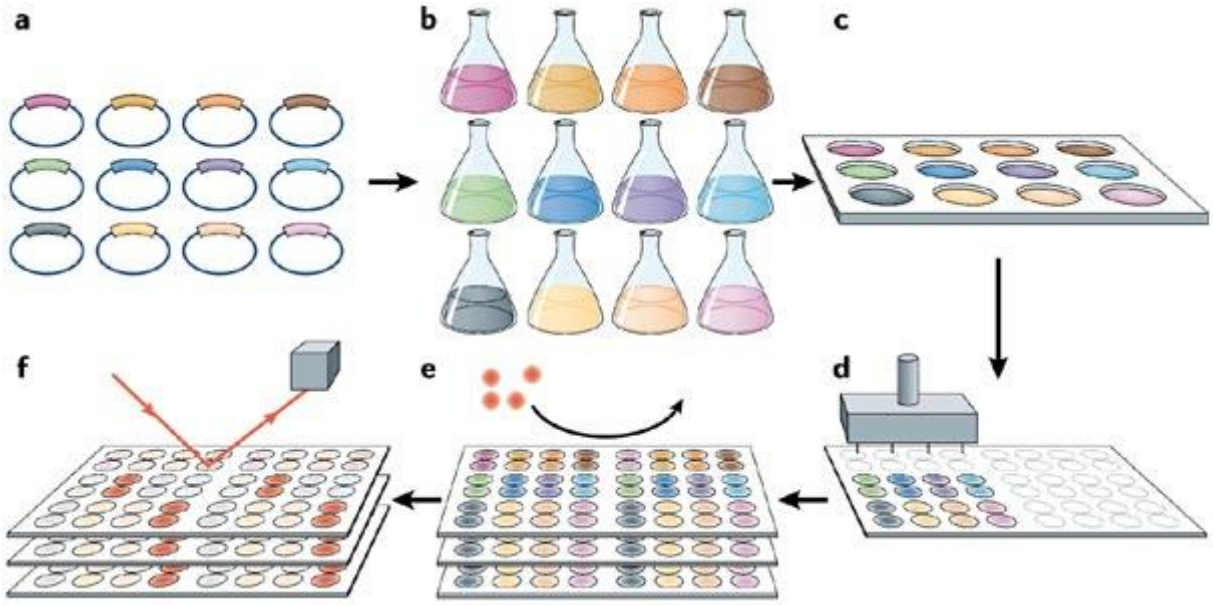


Figure 15: A brief overview of the steps involved in the creation and use of proteome microarrays (Kung and Snyder, 2006). Open reading frames are cloned in a suitable expression vector (a) and transformed in the host organism (b). Proteins are purified (c) and transferred to a high-density chip (d). Subsequently, the chip is assayed with labeled probe, e.g. a fluorescent labeled protein of interest, and unbound probe is washed away (e). Finally, the probe is detected, typically by fluorescence scanning (f).

in the presence of radioactive ATP can reveal phosphorylations (Feilner and Kersten, 2007). In another study, protein arrays were used to test antibody specificity and cross-reactivity (Kersten and Feilner, 2007). Now it is waiting for a protein chip covering a whole plant proteome, a proteome chip, like they already exist for yeast (Zhu et al., 2001). As we are dealing here with an *in vitro* assay, it is recommended to validate identified interactions with a second assay.

Protein display technologies

The best known example of protein display is via filamentous bacteriophages that propagate in *E. coli* (Scott and Smith, 1990; Rodi and Makowski, 1999). A wide variety of ligands (antibodies, peptides, small proteins...) can be expressed by the phage as a fusion with a coat protein, and these are displayed at the surface of the phage particle, where they can be selected through interaction with any given target, e.g. a protein of interest. After selection, the phage-displayed polypeptide can be characterized by amplification and sequencing of the corresponding gene. This is possible due to the physical link between the polypeptide

and its coding gene. Although phage display can be used to identify proteins interacting *in vitro* with given targets, it is not widely used in PPI-screens. This is mainly due because it is limited to the study of small- to medium-sized proteins lacking eukaryotic PTMs. However, the phage display technology is often used to screen for antibodies against a given target by displaying antibody fragments (Benhar, 2007). To overcome the limitations of phage display, new display platforms have been developed, like surface display on yeast (Pepper et al., 2008) or on mammalian cells (Ellmark et al., 2004; Wolkowicz et al., 2005).

1.5.2.2 Mapping functional relationships

Next to methods analyzing real physical interactions between proteins, analysis of functional relationships between genes can provide further insight into biological processes. For example, if two genes lead to a similar phenotype when mutated or have a similar expression pattern, there is a good chance that they act in the same biological pathway or even in the same protein complex. Although deduction of functional relationships is not a direct evidence of interaction, it is often used to assign confidence to observed physical interactions. Here, two methods analyzing functional relationships are briefly discussed.

Genetic interactions

Genetic interactions reflect functional relationships between genes, in which the phenotypic effect of one gene is modified by another. By comparing the effect of mutating each gene individually to the effect of the double mutant, one can identify genetic interactions. In the extreme, synthetic lethality can occur when the combination of two mutations leads to lethality (Beyer et al., 2007). In plants, synthetic lethal genetic interactions have been explored mainly by RNAi to detect genes whose products act in the same essential pathway (Hartung et al., 2006; Yin et al., 2007). In yeast, large networks of genetic interactions are being measured genome-wide through the techniques of synthetic genetic arrays (SGA) and diploid-based synthetic-lethality analysis on microarrays (dSLAM) (Ooi et al., 2003; Ooi et al., 2006; Pan et al., 2007).

mRNA co-expression

Correlations between transcript expression levels of different genes are often inferred from microarray data. These correlations can indicate co-expression and regulatory relations

between genes and their products (von Mering et al., 2005). Clustering methods have been developed to identify similarity of expression of different genes across multiple samples (Gasch and Eisen, 2002). However, co-expression of two genes does not necessarily mean that their products also physically interact. Nevertheless, Gavin and co-workers demonstrated that pairs of proteins within cores or modules tend to be often co-expressed at the same time during the cell cycle and even at a similar copy number (Gavin et al., 2006). So, cores and modules in complexes are often the result of a coordinated gene expression. As a consequence, mRNA co-expression might be used to assign confidence to protein interaction data.

1.5.2.3 Computational prediction of PPIs and *Arabidopsis* PPI databases

As experimental methods do not reach full proteome coverage in identifying PPIs, and as they may be biased towards certain protein types and subcellular localizations, there is a need for computational methods to predict PPIs. These predictions can than be used to evaluate experimentally derived PPIs or to choose potential targets for experimental screenings. Many of these computational methods infer functional relationships between potentially interacting proteins, rather than predicting direct physical associations. To date, several methods are developed to predict PPIs (Shoemaker and Panchenko, 2007). Some are based on the close linkage of genes encoding interacting gene products (gene cluster, gene neighborhood, and Rosetta stone methods), on co-evolution patterns in interacting proteins (sequence co-evolution methods or 'interologs'), and on the co-expression of genes. Some methods screen for certain patterns of co-occurrence in interacting proteins, protein domains, and phenotypes (phylogenetic profiles and synthetic lethality), while others use the presence of sequence/structural motifs characteristic only for interacting proteins (classification or association methods). Yet other tools predict interaction based on topological analysis of gene or protein networks. Next to these prediction methods, text-mining tools were developed to extract gene pairs from biological literature, for example co-occurrence analysis extracts functional relations if a pair of genes appears within the same abstract (von Mering et al., 2007).

Evidence from multiple heterogeneous data sources, both experimental and predicted, is often combined to improve the reliability of relations. For *Arabidopsis*, Li and coworkers

proposed a 2-layered Bayesian network approach for extraction and integration of gene functional relations from diverse biological data sources (Li et al., 2006a). They extracted relations from gene expression, literature and genome sequence and integrated them to score the reliability of the extracted relations. A similar approach was used to build the *Arabidopsis thaliana* protein interactome database AtPID (Cui et al., 2008). In this work they integrate data from ortholog interactome, microarray profiles, GO annotation, conserved domain and genome context together with manually curated data from the literature and from PPI databases, and from enzyme complexes present in KEGG. Another predicted interactome for *Arabidopsis* was build from interacting orthologs in yeast, nematode worm, fruit fly, and human (Geisler-Lee et al., 2007) (Figure 16). However this approach is restricted to conserved interactions and consequently plant-specific interactions are missed. The database containing these 'interologs', together with confirmed interactions from BIND, the Biomolecular Interaction Network Database, and from high-density protein arrays and other literature sources, can be easily queried through the *Arabidopsis* Interaction Viewer. Another interesting relational database, the *Arabidopsis* reactome, is a manually curated database of biological processes, not to confuse with the *Arabidopsis* interologs inferred from human orthologous interactions, present in the Reactome database. Finally, PPIs for *Arabidopsis* are also present on the ftp-server of TAIR and in general PPI databases, like Intact.

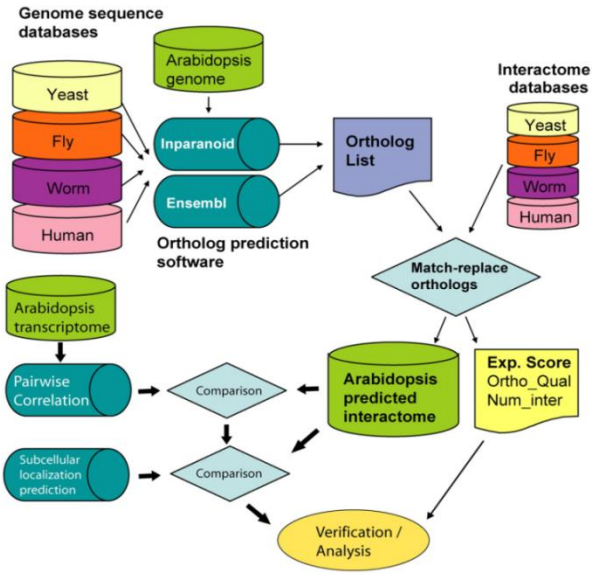


Figure 16: Flowchart for the predicted *Arabidopsis* interactome. *Arabidopsis* orthologs were identified using INPARANOID and ENSEMBL algorithms from genome db of yeast, nematode, fruitfly, and human. If orthologs were found for both partners of a known interaction in the reference species, that interaction was mapped to corresponding *Arabidopsis* genes. This generated the predicted interactome and a confidence value based on the amount of supporting evidence. Subsequent verification and analysis examined each protein pair using co-expression (AtGenExpress) and checked for co-localization using SUBA.

Notes to Chapter 1:

Chapter 2

A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*.

Contributions to chapter 2:

The first 3 authors contributed equally to this paper. The major part of the publication was written by Jelle Van Leene, while Hilde Stals wrote the body of the discussion and Dominique Eeckhout the mass spectrometry-related part. Jelle Van Leene developed the Gateway based high-throughput cloning platform, adapted tandem affinity purification for plant cells, and coordinated the development of the whole technology platform together with Hilde Stals and Geert De Jaeger.

Chapter 2: A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*.

Jelle Van Leene, Hilde Stals, Dominique Eeckhout, Geert Persiau, Eveline Van De Slijke, Gert Van Isterdael, Annelies De Clercq, Eric Bonnet, Kris Laukens, Noor Remmerie, Kim Hendrickx, Thomas De Vijlder, Azmi Abdelkrim, Anne Pharazyn, Harry Van Onckelen, Dirk Inzé, Erwin Witters, and Geert De Jaeger

Abstract

Defining protein complexes is critical to virtually all aspects of cell biology, because many cellular processes are regulated by stable protein complexes and their identification often provides insights into their function. We describe the development and application of a high-throughput tandem affinity purification/mass spectrometry platform for cell suspension cultures to analyze cell cycle-related protein complexes in *Arabidopsis thaliana*. Elucidation of this protein-protein interaction network is essential to fully understand the functional differences between the highly redundant cyclin/cyclin-dependent kinase modules, which are generally accepted to play a central role in cell cycle control, in all eukaryotes. Cell suspension cultures were chosen because they provide an unlimited supply of protein extracts of actively dividing and undifferentiated cells, which is crucial for a systematic study of the cell cycle interactome in the absence of plant development. Here, we report the mapping of a protein interaction network around six known core cell cycle proteins by an integrated approach comprising generic Gateway-based vectors with high cloning flexibility, the fast generation of transgenic suspension cultures, tandem affinity purification adapted for plant cells, matrix-assisted laser desorption ionization tandem mass spectrometry, data analysis, and functional assays. We identified 28 new molecular associations and confirmed 14 previously described interactions. This systemic approach provides new insights into the basic cell cycle control mechanisms and is generally applicable to other pathways in plants.

Manuscript published in: Molecular and Cellular Proteomics 6, 1226-1238 (2007)

2.1 Introduction

Cell division is a fundamental biological process that shares conserved features and controls in all eukaryotes (Nurse, 1994). However, plants have some special features that give control of cell cycle progression a particular importance, and which might be the reason why plants evolved novel molecules orchestrating cell division (Inze and Veylder, 2006). In *Arabidopsis thaliana*, more than 100 core cell cycle genes have been described (Vandepoele et al., 2002; Menges et al., 2005). As in other eukaryotes, CDKs and cyclins govern the plant cell cycle. No less than 29 different CDKs and 52 cyclin-related genes have been identified so far (Menges et al., 2005). However, the molecular interface between the CDK/cyclin complexes, their substrates and interactions with other proteins are, to a large extent, unexplored in plants. Now that the *Arabidopsis* ORFeome is partially being cloned (Hilson, 2006), a next step is the systemic proteomic study to decipher the cell cycle interactome that controls cell division in plants.

In recent years, new technologies have been developed to study protein-protein interactions under near-physiological conditions. Especially tandem affinity purification (TAP) (Rigaut et al., 1999) combined with MS-based protein identification is a powerful approach that has led to the first genome-wide screens for protein complexes in yeast (Gavin et al., 2002; Ho et al., 2002; Gavin et al., 2006; Krogan et al., 2006). TAP strategies have also been developed for transgenic plants (Rubio et al., 2005). However, plants contain only a minor fraction of dividing cells, mostly concentrated in proliferating tissues of the meristems. Moreover, cell cycle proteins are low abundant and some of them have a cell cycle-dependent expression profile. Therefore, cell suspension cultures, rather than complete plants, offer an unlimited supply of protein extracts derived from dividing cells, expressing more than 85% of the core cell cycle regulators (Menges et al., 2005). Furthermore, cell suspensions can be synchronized, by for instance growth factor starvation (Menges and Murray, 2002), making it possible to focus on a specific cell cycle transition (e.g. G1/S or G2/M) to better understand the functional relationship between the different protein complexes during cell cycle progression. Therefore, the *Arabidopsis* cell suspensions are most suited to study the core cell cycle interactome.

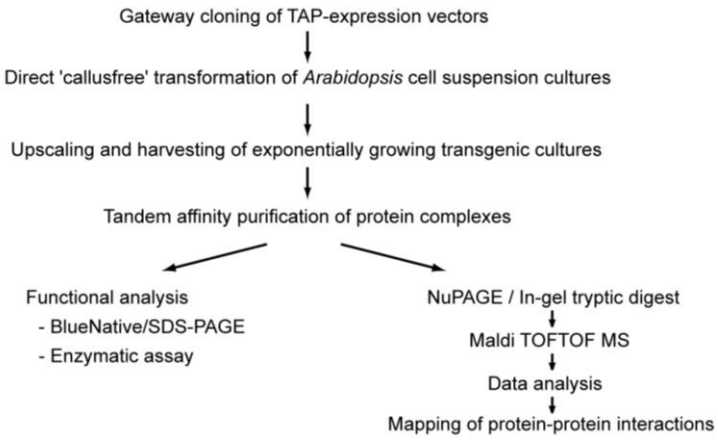
Here, we report on the implementation of a high-throughput TAP platform for *Arabidopsis* cell suspension cultures, combining different technologies to characterize protein-protein interactions in plants. New transformation vectors with high cloning flexibility were designed, allowing cloning of any promoter, ORF, and tag in one step. A streamlined transformation procedure for the fast generation of multiple series of transgenic cell suspension cultures was set up. The original yeast TAP protocol (Rigaut et al., 1999) was adapted for plant suspension cells. Tools were implemented for the high-throughput identification of *Arabidopsis* proteins using MALDI-tandem-MS and data-analysis. Proof of concept for the methodology is shown by protein complex identification for six tagged core cell cycle proteins which were randomly chosen: four different CDKs (CDKA;1, CDKB1;1, CDKD;2, and CDKF;1), and two regulatory subunits (CKS1 and CYCD3;1). Finally, data are provided on the activity of the isolated complexes via a functional assay, and incorporation of the expressed bait protein into physiological complexes is illustrated via Blue Native (BN)/SDS-PAGE.

2.2 Results

2.2.1 Expression of TAP tagged proteins in cell suspension cultures

The strategy to purify and characterize protein complexes from *Arabidopsis* suspension cultures ectopically expressing TAP-fusions is depicted in Figure 1A. To maximize cloning throughput, Gateway-compatible vectors were designed for both N- (pKNTAP) and C-terminal tagging (pKCTAP) (Figure 1B). The vectors give maximal cloning flexibility, because the MultiSite Gateway cassette in pKCTAP enables simultaneous cloning of a promoter sequence, ORF, and a tag of choice. In pKNTAP, the promoter and the ORF can be cloned in one step via independent recombination. We used the TAP tag introduced by Rigaut *et al*., consisting of two IgG-binding domains of the *Staphylococcus aureus* protein A (ZZ) and a CBP, separated by a TEV protease cleavage site.

A.



B.

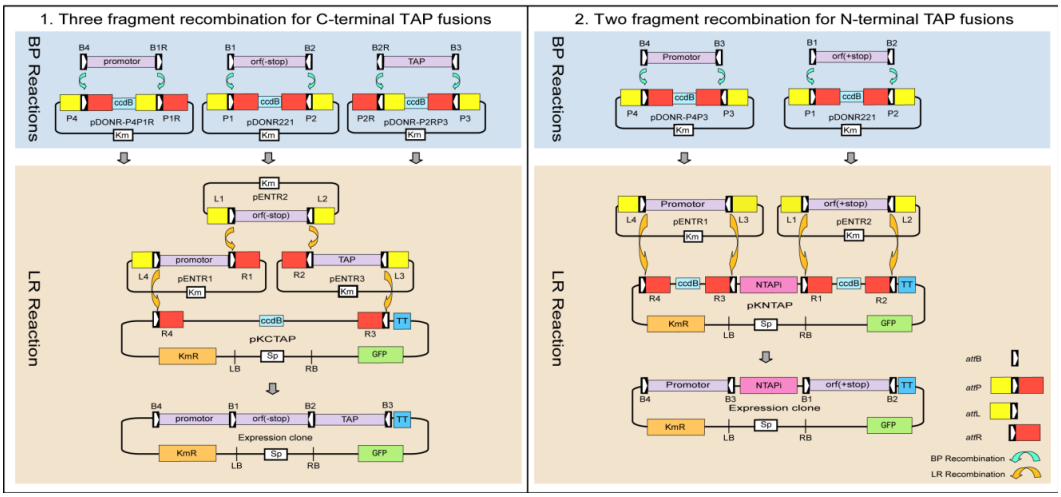


Figure 1. Strategy followed to clone, express, purify and identify TAP tagged proteins and their interacting partners. (A) Summary of the technology platform implemented for screening of cell cycle related protein-protein interactions in plant cells. **(B)** Overview of the TAP-construct cloning strategy. (1) For C-terminal TAP fusions, a three-fragment recombination strategy was used. Entry vectors pENTR1, pENTR2, and pENTR3 were produced in a BP clonase reaction that transferred a PCR amplicon (promoter, ORF without stop codon, and TAP tag, respectively) flanked by the appropriate *att* sites (B4 and B1R, B1 and B2, or B2R and B3) in one of the three compatible donor vectors (pDonrP4P1R, pDonr221 or pDonrP2RP3; http://www.invitrogen.com/). Subsequently, the three fragments were assembled into the pKCTAP destination vector in a single MultiSite LR clonase reaction to produce an expression clone. Besides the Gateway gene cassette, pKCTAP contained between T-DNA border sequences (LB and RB) a kanamycin resistance gene for selection of transformed cells and a GFP expression cassette as visible marker for transformation. (2) For N-terminal TAP fusions, a two-fragment recombination strategy was used. The promoter was cloned by BP clonase reaction in pDonrP4P3, the ORF (plus stop codon) in pDonr221. Subsequently, the two fragments were assembled into the pKNTAP destination vector, which contained the NTAPi-tag (Rohila et al., 2004), during a single MultiSite LR clonase

reaction to produce an expression clone. pKNTAP also comprised a kanamycin resistance gene for selection of transformed cells and a GFP expression cassette as visible marker for transformation. *TT*, CaMV 35S transcription termination sequence; *CcdB*, toxic killer gene for negative selection.

In yeast, the endogenous protein is replaced by the tagged version through homologous recombination in haploid cells (Gavin et al., 2002). However, in higher eukaryotes, the tagged proteins are normally produced in the presence of the untagged endogenous version, which might compete for incorporation into multiprotein complexes. The accumulation level of the tagged protein might thus be an important parameter for complex isolation. Therefore, we first compared the expression levels for five different cell cycle proteins, all tagged C-terminally (CDKA;1, CDKB1;1, CDKD;2, CDKF;1, and CKS1) under control of either a strong constitutive 35S promoter of the cauliflower mosaic virus or their own promoter. The *Agrobacterium*-mediated transformation procedure combined with callus selection was used to generate transgenic *Arabidopsis* cell suspension cultures. Several cultures were made for each construct. Protein extracts of these cultures were screened by immunoblotting with anti-CBP and for each construct the culture with the highest transgene expression was selected for further analysis. As expected, all recombinant proteins accumulated at higher levels when expressed under control of the constitutive promoter compared to the endogenous promoters (Figure 2A). For CDKB1;1, CKS1, and CDKD;2, only very low levels or even no tagged proteins could be detected under control of their endogenous promoter (Figure 2A). Next, we compared the protein levels of the overexpressed fusion proteins, CDKA;1 and CKS1, to the corresponding endogenous proteins. Immunoblot analysis with polyclonal antisera against the two proteins revealed that the recombinant CKS1 concentration in the overexpressing culture was higher than that of the endogenous counterpart (Figure 2B). However, for recombinant CDKA;1 under control of the 35S promoter, only cultures with accumulation levels lower than those of the endogenous CDKA;1 could be obtained (Figure 2C), probably because of gene dosage effects, although a higher turn-over of the TAP-tagged CDKA;1 proteins could not be excluded. Given the high ploidy level (8n) for the *Arabidopsis* cultures we used, the average transgene copy number per cell might be lower than that of the endogenous *Arath;CDKA;1* gene, whereas for CKS1, there might be a compensation by the stronger 35S promoter. Because purifications of

tagged proteins produced under control of a strong constitutive promoter followed by MS provided the best coverage of protein complex detection (see below), we continued with the overexpression strategy.

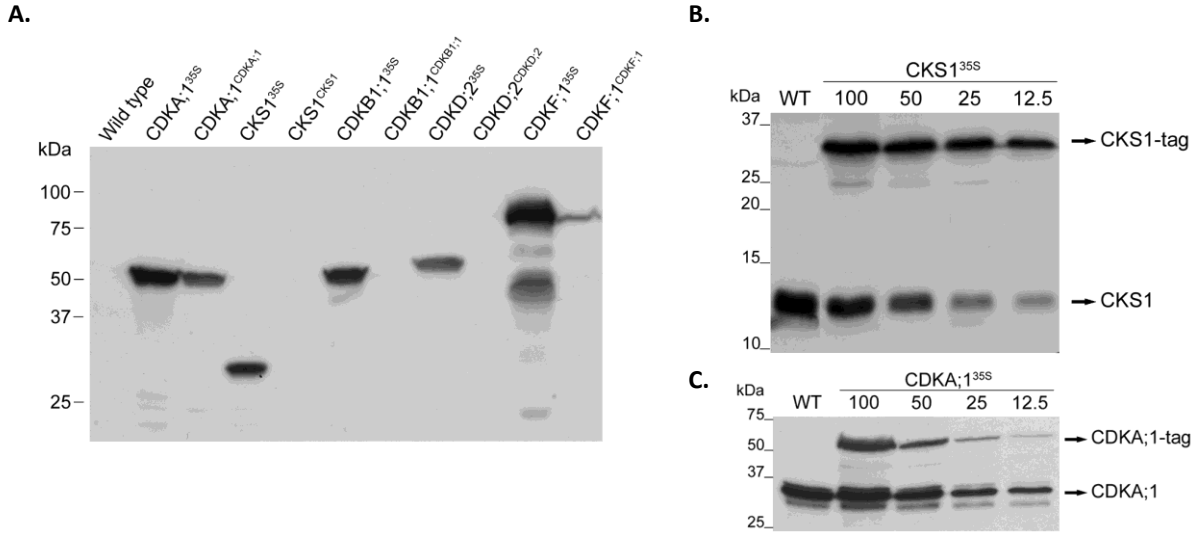


Figure 2. Expression analysis of tagged proteins in transgenic cell suspension cultures. (A) Comparison of transgene expression under control of the constitutive 35S promoter versus the endogenous promoter. Expression was analyzed by immunoblotting with an anti-CBP antibody (1/1000). Total protein extract of 2-day-old cultures (60 µg) was separated by 12% SDS-PAGE and immunoblotted. Expected recombinant molecular masses are 54.6 kDa for CDKA;1, 31.1 kDa for CKS1, 55.9 kDa for CDKB1;1, 59.8 kDa for CDKD;2, and 74.4 kDa for CDKF;1. **(B-C)** Expression analysis of recombinant TAP-tagged CKS1 **(B)** or CDKA;1 **(C)** versus the corresponding endogenous protein by immunoblotting with an anti-CKS1 (1/2500) **(B)** or anti-PSTAIRE (1/2500) **(C)** antiserum. Of total protein from the untransformed cells, 50 µg was analyzed. For the transgenic cultures overexpressing the recombinant protein a series of 100, 50, 25, and 12.5 µg total protein extract were separated. To prevent interaction between IgG present in the used rabbit antisera and ZZ domain of the TAP tag, thereby overestimating accumulation levels of the recombinant protein, human serum was added (1/250) to the blocking buffer during incubation with primary and secondary antibodies.

2.2.2 Optimization of cell suspension transformation

Protocols for *Arabidopsis* cell suspension transformations published so far, are based on *Agrobacterium* cocultivation, followed by transgenic callus selection on agar plates, growth of individual calli, and finally callus resuspension (Forreiter et al., 1997; Menges and Murray, 2004, 2006), the whole procedure taking approximately 3 to 4 months. Especially growing individual calli to sufficient biomass for resuspension is very time consuming and laborious, and hence less suited for a high-throughput setup. Therefore, we changed the

transformation procedure by fine-tuning selection criteria of transformed cells and by eliminating callus selection (see Experimental Procedures). In this manner, transgenic cultures could be obtained in approximately 6 weeks. To assess large-scale utility of the "direct-callus-free" transformation procedure, 20 different randomly chosen cell cycle genes were overexpressed as TAP-tagged fusion proteins in *Arabidopsis* suspension cultures. Transgene expression was analyzed in total protein extracts of exponentially growing cultures by immunoblotting with an antibody against the CBP tag. In all transgenic cultures, the respectively tagged fusion proteins were detected and migrated on a 12% SDS-PAGE at the correct molecular mass (Figure 3A), except for CycD2;1, CycD3;1, E2Fa, SMR2, and SMR3 whose migration was slower than expected. At least for CycD3;1 and E2Fa, this aberrant migration had previously been reported (Magyar et al., 2000; Planchais et al., 2004). For seven proteins, four of which were cyclins, additional protein bands with lower molecular masses were detected, probably due to proteolytic degradation of the unstable proteins. Furthermore, accumulation levels among the 20 different proteins varied largely, despite the overexpression under control of the same 35S promoter. Independent transformations with the same transgene resulted in different cultures with similar expression levels, hence transformation variation could be excluded (Figure 3B).

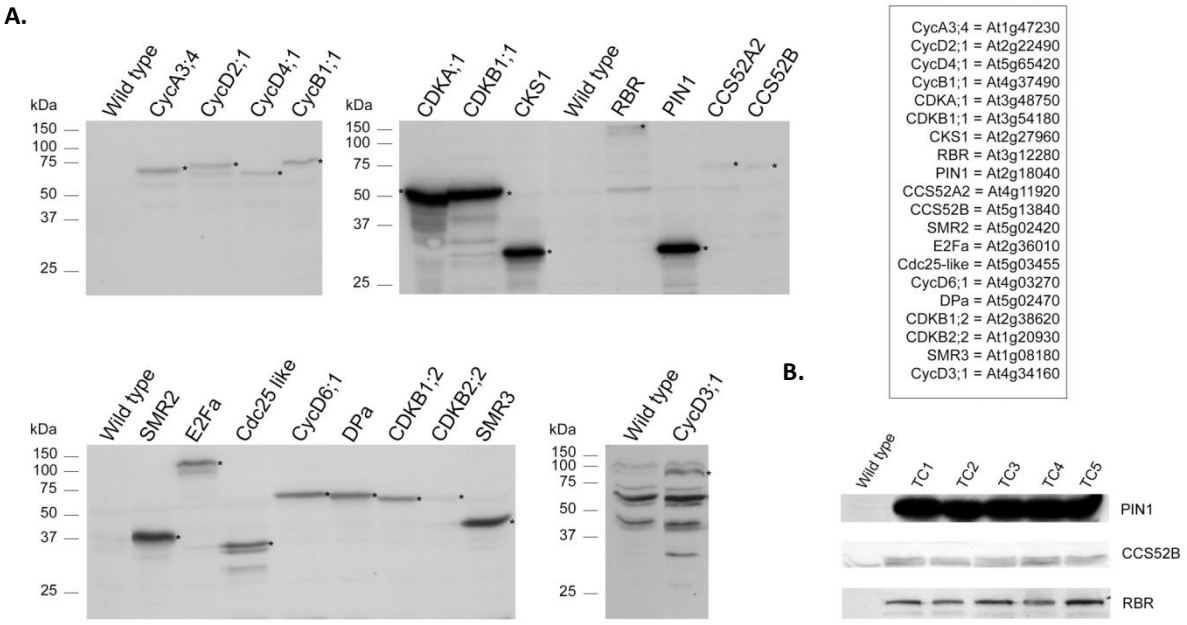


Figure 3. Transgene expression analysis of cultures obtained by transformation without callus selection. (A) Detection of 20 different TAP-tagged proteins (with their corresponding accession numbers) in total protein

extracts of exponentially growing cell suspension cultures. In each lane, 60 µg extract was loaded and recombinant proteins were detected by immunoblotting with an anti-CBP antiserum (1/1000). Bands corresponding to full-length fusion proteins are indicated with an asterisk. **(B)** Evaluation of variation in expression levels due to different transformation events. For three different transgene constructs, five independent transgenic lines were made. Transgene expression was analyzed in 60 µg of total protein extract through immunoblotting with an anti-CBP antiserum. *TC*, transformed culture.

2.2.3 TAP of functionally active protein complexes

Two-dimensional BN/SDS-PAGE was used to analyze the incorporation of tagged proteins into the corresponding physiological complexes. Protein extracts of the cell cultures overproducing C-terminally TAP-tagged CDKA;1 under control of the 35S promoter were separated by two-dimensional BN/SDS-PAGE followed by immunoblot analysis of the bait and the endogenous protein with anti-CDKA;1 polyclonal serum. Recombinant CDKA;1 had a migration pattern similar to that of the endogenous CDKA;1, ranging from non-complexed monomeric CDKA;1 to protein complexes with a molecular mass more than 669 kDa (Figure 4A.1), demonstrating that the tagged protein competes with the endogenous protein to be built into physiological protein complexes.

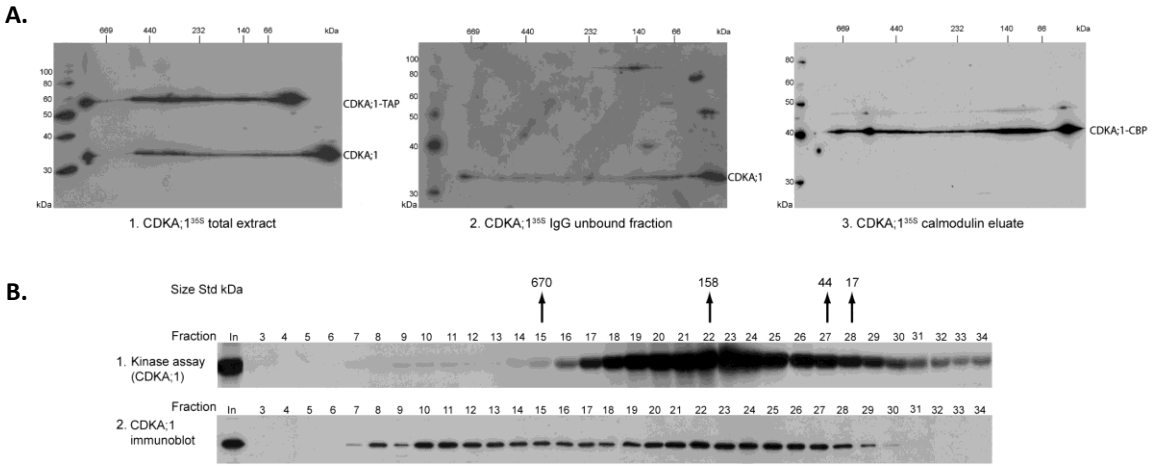


Figure 4. Protein complex analysis by BN/SDS-PAGE (A) and size exclusion chromatography (B). (A) Protein complexes fractionized in the first dimension by BN-PAGE. In the second dimension, proteins were separated by SDS-PAGE and gels were immunoblotted with the anti-CDKA;1 antiserum (1/5000). For CDKA;1^{35S}, 400 µg of protein was analyzed from the total protein extract (*1*) and the IgG-unbound fraction (*2*), and 55 µg from the final calmodulin eluate (*3*). Recombinant CDKA;1 and CDKA;1 fused to the complete TAP tag and the CBP tag are indicated as CDK1;1-TAP and CDKA;1-CBP, respectively. **(B)** CDKA;1 TAP eluate (2 mL) fractionated on a Superdex 200 (300/10) size-exclusion chromatography column. From each 500 µL fraction, 25 µL was assayed

for kinase activity with histone H1 as a substrate (*1*), and 10 µL was analyzed for the presence of the bait protein by immunoblotting with an antiserum against CDKA;1 (1/5000) (*2*). Arrowheads indicate the elution positions of marker proteins with their molecular masses (kDa).

The original TAP procedure from yeast (Rigaut et al., 1999) was adapted to simultaneously purify TAP-tagged protein complexes from different plant cell cultures. Major changes were decreasing the incubation time with both affinity chromatography matrices, adding protease inhibitors to all buffers, thereby reducing protein degradation of unstable cell cycle proteins, and increasing the EGTA concentration during the calmodulin-agarose elution. Typically, 15 g of frozen cell suspension material was used to prepare a protein extract containing 200 to 300 mg total protein. In a first affinity purification step, extracts were incubated with IgG resin and bound complexes were eluted by tag cleavage with TEV protease. Co-eluting non-interacting proteins and the TEV protease were further separated from the tagged proteins and their associations in the flow through of the second affinity step. The bait and interacting proteins were finally eluted from the calmodulin agarose via EGTA-mediated removal of calcium. Two dimensional BN/SDS-PAGE followed by immunoblot analysis was used to evaluate the purification procedure. Different steps of the purification of the culture overproducing TAP-tagged CDKA;1, were analyzed in that manner. In the flow-through fraction of the first affinity chromatography step, only protein complexes with the endogenous CDKA;1 were detected (Figure 4A.2), demonstrating that the IgG resin bound all protein complexes containing the recombinant CDKA;1. In the final calmodulin eluate, the recombinant CDKA;1-containing protein complexes had a migration pattern similar to that of the starting material (Figure 4A.1,3), indicating that the TAP protocol allowed the isolation of intact CDK protein complexes.

To further analyze the functional activity of the purified complexes and thus their integrity, the CDKA;1 calmodulin eluate was fractionized by size-exclusion chromatography. All fractions were analyzed for the presence of the bait protein by immunoblotting with an anti-CDKA;1 antibody and assayed for CDK activity with histone H1 as a substrate (Figure 4B). The maximum of the histone H1 kinase activity migrated around 100 kDa (fraction number 23), and was associated with a maximum for CDKA;1 fusion protein. These observations are consistent with previously reported results in tobacco Bright Yellow-2 cells (Porceddu et al.,

2001), and demonstrate that the overexpressed CDKA;1 fusion protein is incorporated into physiologically active complexes, which can be isolated through our TAP protocol. Immunoblot analysis of the sized fractions also showed that besides the active complex, the recombinant CDKA;1 protein is part of inactive high-molecular-mass complexes (Figure 4B). These data are consistent with the two-dimensional BN/SDS-PAGE results.

2.2.4 Identification of co-purifying proteins by MS

To identify protein interactions for the cell cycle baits CDKA;1, CDKB1;1, CKS1, CDKD;2, CDKF;1, and CYCD3;1, TAP purifications were performed on cultures of C-terminally tagged CDKA;1, CDKB1;1, CKS1, CDKD;2, CDKF;1, and N-terminally tagged CYCD3;1 under control of the 35S promoter, and cultures of C-terminally tagged CDKA;1 and CDKF;1 under control of the endogenous promoter. Each culture was at least purified twice independently. Eluted proteins were separated by 4-12% NuPAGE and visualized by colloidal Coomassie staining. Mostly, the tagged protein was detected as the most prominent band on the gel (Figure 5).

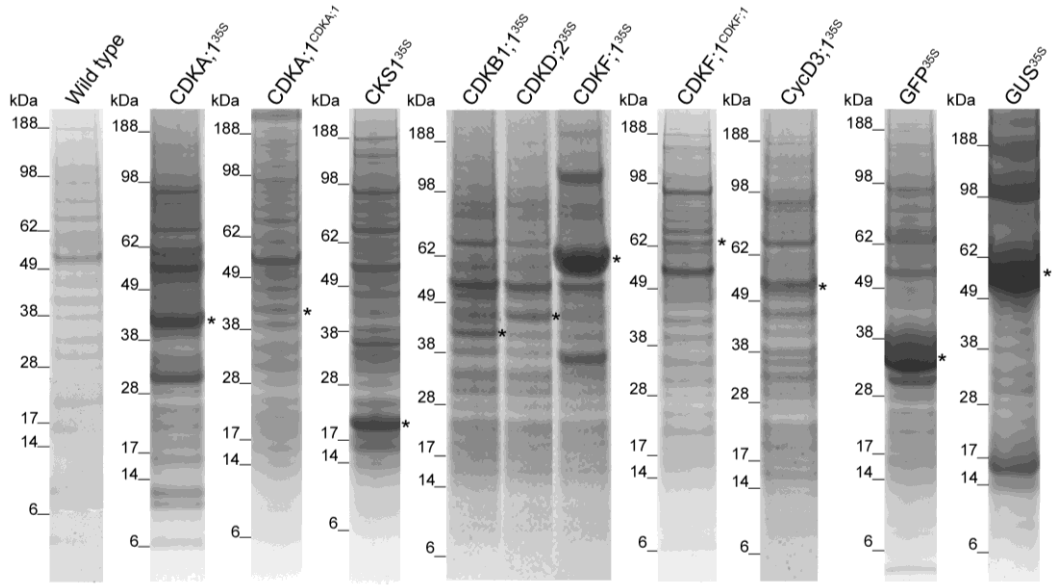


Figure 5. Analysis of the TAP protein eluates. TAP-tagged protein complexes were purified from transgenic plant cell suspension cultures, precipitated with TCA (25% v/v), separated by 4-12% Nu-PAGE, and visualized with colloidal Coomassie G-250 staining. Bait proteins are indicated with an asterisk.

Protein bands were excised, subjected to tryptic digestion, and analyzed via MALDI-TOF-TOF-MS. The resulting peptides were assigned to specific proteins with the in-house SNAPS database (http://www.ptools.ua.ac.be). This non-redundant database combines all publicly available *Arabidopsis* protein sequences. By querying the complete set of available

sequences at a time, the best possible hits were retrieved. Contaminating proteins due to experimental background were determined by purifications on wild-type and two transgenic cultures overexpressing TAP tagged fusions of heterologous proteins, β-glucuronidase (GUS) and green fluorescent protein (GFP) (Figure 5). Most contaminants are high abundant proteins, such as chaperones, cytoskeleton proteins, ribosomal proteins, metabolic enzymes, or protein translation factors (Supplemental Table 8). Identical or similar proteins were found as common contaminants in other protein-protein interaction studies (Gavin et al., 2002; Ho et al., 2002; Shevchenko et al., 2002; Archambault et al., 2004; Bouwmeester et al., 2004; Rohila et al., 2006). Therefore, to increase the stringency of the data set, these proteins were systematically subtracted from the lists of purified proteins. The remaining identified proteins were divided into two groups: 43 proteins that could be confirmed experimentally (Table I) and 186 proteins that were identified only once per bait (Supplemental Table 9).

Table I: Proteins identified by MALDI-TOF-TOF-MS after a TAP procedure from cultures producing CDKA;1 and CDKF;1 under control of their endogenous promoter, and CDKA;1, CDKB1;1, CKS1, CDKD;2, CDKF;1, and CYCD3;1 under control of the 35S-CaMV promoter. Proteins co-purifying with the bait proteins were shown only when they were confirmed in more than one experimental repeat. The GO annotations or references related to cell cycle control mechanisms are given as GO ID/term or reference.

Accession number	Protein name	GO-ID[a] or reference (*number*)
CDKA;1^{35S}		
At2g27960	CKS1	GO: 74, 278, 4693, 42023
At2g27970	CKS2	GO: 7049, 4693
At2g20580	26S proteasome regulatory subunit S2 (RPN1)	GO: 74
At5g23540	26S proteasome regulatory subunit, putative	(*54*)
At1g64520	26S proteasome regulatory subunit, putative (RPN12)	(*55*)
At2g22490	CYCD2;1	GO: 74, 80, 16538
At5g40460	Hypothetical protein	
At3g19150	KRP6	GO: 4861, 45736, 30332
At1g23190	Phosphoglucomutase, cytoplasmic, putative	
At4g28470	26S proteasome regulatory subunit, putative	

At5g63610	CDKE;1	(*3*)
At5g65420	CYCD4;1	GO: 80, 74, 16538
At5g10440	CYCD4;2	GO: 74, 4693
At1g10690	Expressed protein	
At2g32710	KRP4	GO: 45736, 4861, 30332, 45786
At3g49240	Pentatricopeptide (PPR) repeat-containing protein	GO: 9793
At2g28000	RuBisCO subunit binding-protein alpha subunit, chloroplast	
At3g17020	Universal stress protein (USP) family protein	
At1g78900	Vacuolar ATP synthase catalytic subunit A	GO: 9555
CDKA;1$^{pCDKA;1}$		
At3g55000	Tonneau 1a	
CDKB1;1^{35S}		
At2g27970	CKS2	GO: 7049, 4693
At2g28000	RuBisCO subunit binding-protein α subunit, chloroplast	GO: 9790
At1g64520	26S proteasome regulatory subunit, putative (RPN12)	(*55*)
CDKD;2^{35S}		
At5g27620	CYCH;1	GO: 74, 4693
At5g08690/At5g08670	ATP synthase beta chain, mitochondrial	
At4g30820	CAK assembly factor-related protein (Mat1)	
At4g16143	Importin α-2, putative (IMPA-2)	
CDKF;1$^{pCDKF;1}$		
At3g16270	Expressed protein similar to cyclin G-associated kinase (O14976)	GO: 74
CDKF;1^{35S}		
At1g67580	CDKG;2	(*3*)
At1g66750	CDKD;2	GO: 79
CKS;1^{35S}		
At3g48750	CDKA;1	GO: 8284, 9574, 4693
At3g54180	CDKB1;1	GO: 4693
At1g47230	CYCA3;4	GO: 74, 16538
At4g14310	Peroxisomal membrane protein-related	
At1g20930	CDKB2;2	GO: 87, 4693, 7346
At2g22490	CYCD2;1	GO: 80, 74, 16538
Arath05g16630	unknown	

At3g53880	Aldo/keto reductase family protein similar to chalcone reductase	
At1g76540	CDKB2;1	GO: 307, 86, 4693
CYCD3;1^{35S}		
At3g48750	CDKA;1	GO: 8284, 9574, 4693
At2g27970	CKS2	GO: 7049, 4693
At5g02220	Expressed protein similar to SIAMESE	(40)
At3g19150	KRP6	GO: 4861, 45736, 30332

[a] 74, Regulation of progression through cell cycle; 79, Regulation of cyclin-dependent protein kinase activity; 80, G1 phase of mitotic cell cycle; 86, G2/M transition of mitotic cell cycle; 87, M phase of mitotic cell cycle; 278, Mitotic cell cycle; 307, Cyclin-dependent protein kinase holo-enzyme complex; 4693, Cyclin-dependent protein kinase activity; 4861, Cyclin-dependent protein kinase inhibitor activity; 7049, Cell cycle; 7346, Regulation of progression through the mitotic cell cycle; 8284, Positive regulation of cell proliferation; 9555, Male gametophyte development; 9574, Preprophase band; 9790, Embryonic development; 9793, Embryonic development; 16538, Cyclin-dependent protein kinase regulator activity; 30332, Cyclin binding; 42023, DNA endoreduplication; 45736, Negative regulation of cyclin-dependent protein kinase activity; 45786, Negative regulation of progression through the cell cycle.

The data set of confirmed interactors in Table I shows enrichment for proteins that are annotated as regulators of cell cycle progression (GO-ID number; Table I). This enrichment was statistically confirmed with the BiNGO tool, implemented as a plug-in for Cytoscape (Maere et al., 2005). BiNGO analysis of the total list of interacting proteins (Table I and Supplemental Table 9) revealed a statistically significant enrichment for proteins involved in regulation of the cell cycle (corrected *p*-value 2.66E-11; Supplemental Figure 2a). When subtracting proteins that could not be identified in at least two experimental repeats, the *p*-value decreased even more (corrected *p*-value 1.45E-17; Supplemental Figure 2b), suggesting that the non-confirmed data set contained proteins unrelated to cell cycle control and occurring occasionally and randomly due to experimental background. Nevertheless, BiNGO analysis on the subset of non-confirmed proteins clustered CKS1, CDKA;1, and two DNA-mismatch repair proteins (MSH6-1 and MLH1) together as involved in DNA-dependent DNA replication (corrected *p*-value 2.14E-2; Supplemental Figure 2c). Moreover, MSH6-1 and MLH1 co-purified with CDKA;1 and CKS1, suggesting that they interacted, albeit without confirmation. Therefore, the non-confirmed identifications have to be evaluated with caution, because they can represent weaker or more transitory associations between

proteins or protein complexes. Especially those that are common between different baits are certainly interesting interactors to be followed up.

Depending on the expression level of the TAP-tagged fusion proteins, the number of identified protein interactions differs a lot. Under control of its endogenous promoter, we could only confirm TONNEAU 1b (TON1b) as an interacting protein, this in contrast to 19 proteins identified after purifications of the overexpressed CDKA;1 bait (Table I). TON1b is a protein involved in cortical microtubule organization (Traas et al., 1995) and is similar to the human FGFR1 oncogene partner (FOP), a protein implicated in centrosomal microtubule anchoring (Yan et al., 2006). Besides the fact that CDKA;1 is known to co-localize with microtubules of the preprophase band (Stals et al., 1997) and this plant-specific cytoskeleton structure is absent in the *ton1* mutant, there is no known correlation between the two proteins. Among the 19 interacting proteins identified after overproducing CDKA;1-TAP, six proteins had been previously reported as CDKA;1 interactors in plants: three D-type cyclins, CKS1, and two KRPs (De Veylder et al., 1997b; De Veylder et al., 2001; Healy et al., 2001; Kono et al., 2003; Kono et al., 2006; Nakai et al., 2006). Reciprocal purifications with CKS1 as bait validated this result and confirmed previously reported associations of CKS1 with A- and B-type CDKs (Boudolf et al., 2001). Besides CKS1, we also identified its close homolog CKS2 as an interacting protein of CDKA;1. CKS2 was also present in the purified fractions of overexpressed CDKB1;1 and CYCD3;1, and in cells expressing CDKA;1 under control of its endogenous promoter (Table I and Supplemental Table 9). In addition to KRP6 and CKS2, CDKA;1 was also identified as interacting protein of the CYCD3;1-TAP fusion. Despite the isolation of KRP6 in the CDKA;1 data set, we could not identify CYCD3;1 in this complex (Table I), probably because of the highly unstable character of this specific D-type cyclin (Planchais et al., 2004). We could also identify an expressed protein, similar to SIAMESE (SIM) as an interactor for CycD3;1 (Table I). SIM is a plant-specific cell cycle regulator that controls endoreduplication onset in *Arabidopsis thaliana* (Churchman et al., 2006). It encodes a nuclear localized 14-kDa protein containing a cyclin-binding motif and a motif found in KRP cell cycle inhibitors. Furthermore, it was found to associate with D-type cyclins and CDKA;1 (Churchman et al., 2006).

In the set of six tagged cell cycle proteins, we also analyzed the CDK-activating kinases (CAKs), CDKD;2 and CDKF;1 that have recently been proposed to represent two types of CAKs in *Arabidopsis*, playing a major role in phosphorylation of CDKs and the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Shimotohno et al., 2006), respectively. We show that overexpressed CDKD;2 fusion protein forms a stable trimeric complex with the regulatory subunit CYCH;1 and the assembly factor MAT1 (Table I). Both proteins have been previously reported as interactors (Rohila et al., 2006; Shimotohno et al., 2006). In mammals and rice (*Oryza sativa*), the trimeric CAK complex is part of a protein complex that forms the general transcription factor TFIIH and is responsible for the CTD-phosphorylation (Lee et al., 2005; Rohila et al., 2006). This complex could not be identified, in contrast to XPD (UVH6), a helicase that links the trimeric CAK to THIIH complex, albeit without confirmation (Supplemental Table 9) and another interactor of CDKD;2, importin α-2 (Table I). The human homolog of importin α-2 (importin α/β) has recently been described in mammalian cells as a nuclear import receptor that binds one of the nuclear localization signal of cyclin H, thereby regulating the nuclear translocation of the cdk7/cyclin H complex (Krempler et al., 2005). Our data show for the first time that a similar regulatory mechanism might be active in plant cells.

CDKF;1 is a known CAK-activating kinase (CAKAK) that phosphorylates other CAKs in a cyclin H-independent manner (Shimotohno et al., 2004). CDKD;2 was co-purified from extracts overproducing the CDKF;1-tagged protein (Table I), demonstrating that the method allows us to purify the physiologically active CAKAK and its associated substrate. However, expression of CDKF;1 under control of its own promoter did not yield CDKD;2 (Table I), instead a expressed protein with similarity to the human cyclin G-associated kinase was identified. In the cell suspension cultures overexpressing CDKF;1, we also identified CDKG;2 as a new interacting protein. Together with its close homolog CDKG;1, CDKG;2 has recently been described as a new member of the CDK family in *Arabidopsis* (Menges et al., 2005) that are characterized by a PLTSLRE motif and have the highest sequence homology to the human galactosyltransferase-associated (GTA) protein kinase p58/GTA. p58/GTA is a member of a $p34^{cdc2}$-related kinase subfamily that interacts with cyclin D3 (Zhang et al., 2002) and might act as a negative regulator of cell cycle progression in mammalian cells (Bunnell et al., 1990).

The human p58/GTA isoform is specifically induced during G2/M phase of the cell cycle and is post-translationally regulated by phosphorylation (Beausoleil et al., 2004). Sequence comparison of the plant CDKG;2 revealed that the activating Thr phosphorylation site in the T-loop, known to be regulated by CAKs, is conserved.

Finally, we mapped the identified interactions around the six baits (Figure 6) with the Cytoscape software. The interactions are depicted as arrows from the baits to the identified associations, both represented as respectively orange and blue nodes. The mapping revealed an extensive network around CDKA;1, the primary CDK for cell division, including its known cyclin partners, CDK inhibitors (KRP4 and KRP6), and the small CDK-binding subunits CKS1 and CKS2, as well as new interactions. In various cases, the detected interactions were bridged by intermediary partners, implying that many interactions, such as CKS1-cyclin or KRP6-CYCD3;1 occur through the CDK subunit.

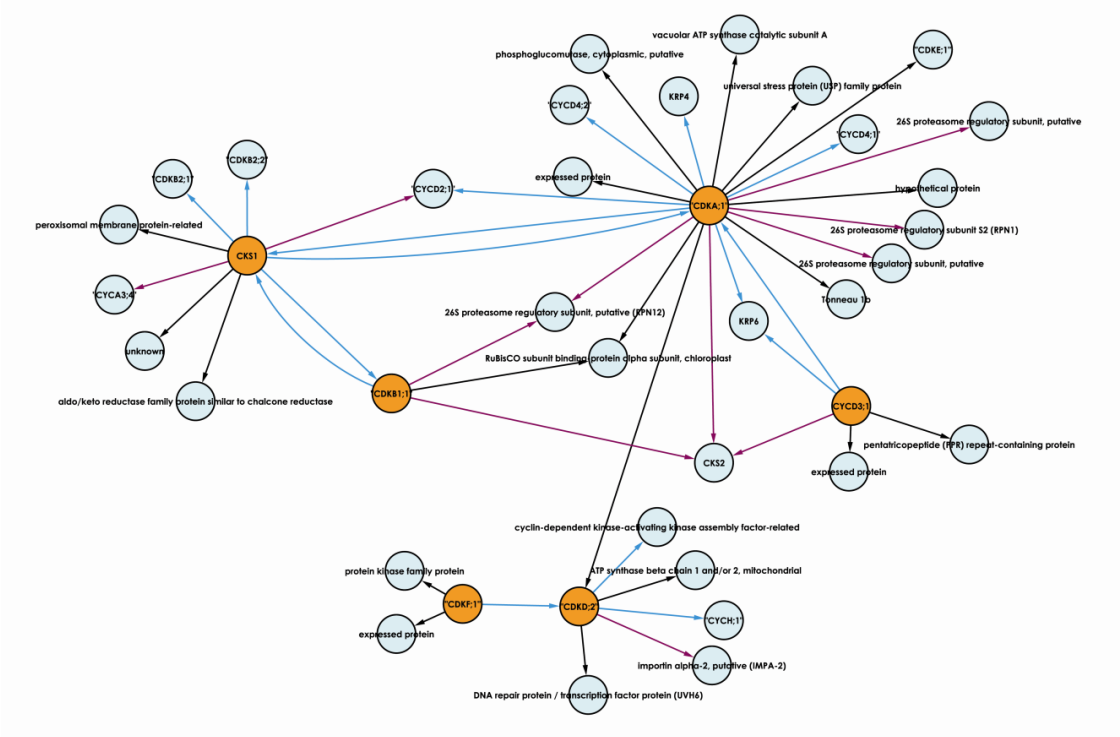


Figure 6. Interactome map for the protein interactions (blue nodes) detected with six core cell cycle proteins (orange nodes). Interactions are represented as arrows, pointing from the bait to the interactor. Known interactions, links previously described in the literature in orthologous systems, and unknown interactions are represented in blue, purple, and black, respectively.

2.3 Discussion

In eukaryotes, cell division is controlled by a large set of genes whose expression and function are both transcriptionally and post-transcriptionally regulated. While DNA arrays continue to provide information regarding changes at the mRNA level, posttranslational modifications, changes in cellular localization or interactions within complexes can only be assessed at the protein level. Here, we report the successful setup of a TAP technology platform that allows us to unravel protein-protein interaction networks in plants. The methodology is based on an integrated approach comprising the fast generation of transgenic cultures overproducing tagged fusion proteins, TAP adapted for plant cells, high-throughput protein identification by tandem-MS, data analysis, and functional assays.

Our generic MultiSite Gateway-based transformation vectors enable high-throughput cloning of large sets of TAP-tagged fusions, with as only time-limiting factor the cloning of the ORF of the genes of interest. In contrast to formerly published TAP vectors for plants (Rohila et al., 2004; Rubio et al., 2005; Rohila et al., 2006), we are able to construct fusions of any ORF with any tag under control of any promoter in one cloning step, as long as they are cloned in the proper Gateway donor vectors. Comparing accumulation levels of a set of randomly chosen tagged cell cycle proteins revealed a large variation despite overexpression under control of the same constitutive promoter. Possible explanations could be either counter selection of transgenic cells expressing aberrant levels of essential cell cycle genes or could be dependence on the nature of the proteins and on posttranslational regulation controlling the abundance of the corresponding endogenous protein. One could expect that relatively unstable proteins, such as cyclins, had lower accumulation levels than more stable and constitutive proteins, such as CDKA;1 and CKS1. The level at which the tagged bait accumulates has a major impact on the yield of purified complexes, and hence the identified interactors. Data sets of interactions from ectopically expressed CDKA;1 and CDKF;1 under a strong constitutive or the endogenous promoter clearly showed this prominent difference. In our system, the tagged protein is produced in the presence of the endogenous protein that might contend for incorporation into multiprotein complexes. Given the high ploidy level of the *Arabidopsis* cell suspension culture used, only overexpression under control of the strong 35S promoter resulted in sufficient amounts of bait protein to compete with the

endogenous protein. The use of an overexpression system probably also facilitates detection of weaker or more transitory associations between proteins. Indeed, cyclins that have a cell cycle-dependent protein accumulation level controlled via ubiquitin/proteasome–dependent degradation could only be isolated when the bait was overproduced. Additional factors determining the number of possible identified protein interactions depends on the nature of the bait itself. For example, the number of interactions for CDKA;1, which is a kind of hub within the cell cycle interactome and is generally accepted to play a central role in cell cycle control, is much higher compared to proteins that have a more specific function. This observation has also been reported in other studies (Gavin et al., 2002; Ho et al., 2002). The likelihood of identifying protein associations with a cell cycle-dependent expression profile might also depend on the harvest time of the suspension culture. In this study, we worked with non-synchronized exponentially growing cells. Flow cytometric analysis of the harvested cell material showed an equal G1/G2 phase distribution (data not shown). Hence, proteins with a cell cycle phase-specific expression pattern will yield fewer interactors, as observed in our data set for CDKB1;1 that has a known G2/M-specific kinase activity and expression profile (Porceddu et al., 2001). However, the amount of purified protein complexes does not solely depend on the expression profile of the transgene, but also on the accessibility and integrity of the TAP tag, and its possible sterical hindrance during complex formation. Analysis by BN/SDS-PAGE of the different purification steps showed that the majority of the CDKB1;1-associated complexes bound only weakly to the first affinity resin (data not shown).

Gel filtration combined with kinase assays was integrated into the platform to analyze the physiological activity of the tagged proteins, thereby proving that our method allows the isolation of active CDK complexes. The activity peaks around 100 kDa, suggesting that these complexes probably consist of CDK and their regulatory subunit, cyclin. Besides the active complexes, CDKA;1-fusion protein was incorporated into large, inactive complexes (>600 kDa). Identification of 26S proteasome regulatory subunits in the TAP eluates suggest that part of these large complexes might be composed of CDKA;1 associating with the 19S regulatory particle of the proteasome, and possibly reflect the targeting of interacting proteins, such as CDK inhibitors or cyclins, to the degradation pathway. The ability of the ubiquitin/proteasome

pathway to selectively degrade a single subunit of a multisubunit complex is a known regulatory switch, including those involving cyclins and CDK inhibitors. Mitotic CDK is inactivated at the end of mitosis by selective degradation of tightly bound mitotic cyclin (Peters, 1998; Criqui et al., 2001), and S phase CDK/cyclin is activated *in vivo* at the G1/S transition upon degradation of its CDK inhibitor (Weinl et al., 2005).

In conclusion, for the set of six different cell cycle proteins, we have identified and confirmed 42 protein-protein associations of which 28 were new interactions, demonstrating that the integrated TAP-MS platform that we developed for plant cells is a powerful tool to systematically identify cell cycle complexes and link proteins of unknown function to signaling pathways during cell division. Extension of the proteomic approach described in this study with the analysis of synchronized transgenic cultures should make it possible to investigate the variations in protein associations during cell cycle progression. Characterizing the global topology and dynamic features of the cell cycle interactome will be a major challenge for the future that certainly will provide further insights into developmental mechanisms.

2.4 Experimental procedures

2.4.1 Vector construction for C-terminal tagging

pKCTAP was constructed by cloning the 35S transcription termination signal of the cauliflower mosaic virus, and a green-fluorescent protein (GFP) expression cassette, recovered from pK7WG2D, into pKm43GW (Karimi et al., 2005).

2.4.2 Vector construction for N-terminal tagging

KNTAP was made by cloning three fragments separately in pUC19SX (Eeckhout et al., 2004) between the appropriate restriction sites: a fragment containing the *att*R4/*att*R3 Gateway (Invitrogen, Carlsbad, CA) cassette amplified from pKCTAP, a fragment containing the NTAPi sequence preceded by the omega leader and Kozak sequence amplified from the vector ntapi.289.gw.gck (Rohila et al., 2004), and a fragment containing the *att*R1/*att*R2 Gateway cassette amplified from pKGW (Karimi et al., 2005). The cloned fragments were checked by DNA sequence analysis and aligned in pK^{-} (M. Karimi, personal communication), together

with a fragment containing the 35S transcription termination signal and the GFP expression cassette cut from pKCTAP.

All entry and destination vectors were checked by sequence analysis. Expression vectors were obtained by MultiSite LR reaction (Gateway) and transformed to *Agrobacterium tumefaciens* strain C58C1Rif^R (pMP90) by electroporation. Transformed bacteria were selected on yeast extract broth (YEB) plates containing 100 µg/mL rifampicin, 40 µg/mL gentamicin, and 100 µg/mL spectinomycin.

2.4.3 Cell suspension cultivation

Wild-type and transgenic *Arabidopsis thaliana* cell suspension cultures ecotype *Landsberg erecta* (PSB-D) were maintained in 50 mL MSMO medium (4.43 g/L MSMO [Sigma-Aldrich], 30 g/L sucrose, 0.5 mg/L α-naphthaleneacetic acid, 0.05 mg/L kinetin, pH 5.7, adjusted with 1 M KOH) at 25°C in the dark, by gentle agitation (130 rpm). Every 7 days, the cells were subcultured in fresh medium at a 1/10 dilution.

2.4.4 Cell culture transformation with callus selection

The *Arabidopsis* culture was transformed by *Agrobacterium* co-cultivation as described previously (Forreiter et al., 1997), with minor modifications. The *Agrobacterium* culture exponentially growing in YEB medium (OD_{600} between 1.0 and 1.5) was washed three times by centrifugation (10 min at 3050*g*) with an equal volume MSMO medium and resuspended in cell suspension-growing medium until an OD_{600} of 1.0. Two days after subcultivation, 3 mL suspension culture was incubated with 200 µL washed agrobacteria and 200 µM acetoseringone, for 48 h in the dark at 25°C with gentle agitation (130 rpm). Plant cells were washed 3 times with 50 mL MSMO containing a mix of three antibiotics (50 µg/mL kanamycin, 500 µg/mL carbenicellin, and 500 µg/mL vancomycin). Cells were plated on MSMO agar (0.8%) containing the antibiotics mix, and stored at 25°C in the dark until callus formation was observed. After 2 weeks, stably transformed calli strongly expressing the fluorescent marker GFP were transferred to fresh MSMO agar plates containing 50 µg/mL kanamycin and further grown for 4-6 weeks. Sixty milligrams of callus tissue was suspended in 10 mL MSMO containing 50 µg/mL kanamycin. After 1 week, 40 mL fresh MSMO was added, and cultures were maintained as described previously. Two weeks later, transgene

expression was analyzed in a total protein extract derived from exponentially growing cells, harvested 2 days after subculturing, by immunoblotting using the anti-calmodulin-binding peptide (CBP) antiserum (1/1000; Upstate, Bedford, MA). Cultures with high transgene expression were gradually scaled up to 200 mL (dilution ratio 1:5), and 2x1 L (dilution ratio 1:10) in 500-mL and 2-L erlenmeyers, respectively. Finally, cell material from 10x 1-L cultures (dilution ratio 1:5) grown for 2 days, was pooled, harvested on a sintered glass filter in liquid nitrogen, stored at -80°C, and processed later for protein extraction and TAP.

2.4.5 Cell culture transformation without callus selection

Agrobacteria were co-cultivated with PSB-D suspension cells under the same conditions as described above. Two days after co-cultivation, 7 mL MSMO containing a mix of three antibiotics (25 µg/mL kanamycin, 500 µg/mL carbenicellin, and 500 µg/mL vancomycin) was added to the cell cultures and grown further in suspension under standard conditions (25°C, 130 rpm, and continuous darkness). The stable transgenic cultures were selected by sequentional dilution in a 1:5 and 1:10 ratio in 50 mL fresh antibiotics mix-containing MSMO medium 11 and 18 days post co-cultivation, respectively. After counterselecting the bacteria, the transgenic plant cells were subcultured weekly in a 1:5 ratio in 50 mL MSMO medium containing 25 µg/mL kanamycin for 2 more weeks. Thereafter, the cultures were grown and scaled up as described above.

2.4.6 Protein extract preparation

Plant material (15 g) of exponentially growing cell cultures, harvested 2 days after subculturing, was ground to homogeneity in liquid nitrogen. Crude protein extracts were prepared in an equal volume (w/v) of extraction buffer (25 mM Tris-HCl, pH 7.6, 15 mM $MgCl_2$, 5 mM EGTA, 150 mM NaCl, 15 mM *p*-nitrophenylphosphate, 60 mM β-glycerophosphate, 0.1% [v/v] Nonidet P-40 (NP-40), 0.1 mM sodium vanadate, 1 mM NaF, 1 mM DTT, 1 mM PMSF, 10 µg/mL leupeptin, 10 µg/mL aprotinin, 5 µg/mL antipain, 5 µg/mL chymostatin, 5 µg/mL pepstatin, 10 µg/mL soybean trypsin inhibitor, 0.1 mM benzamidine, 1 µM *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E64), 5% [v/v] ethylene glycol) with an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC) at 4°C. The soluble protein

fraction was obtained by a two-step centrifugation at 36,900*g* for 20 min and at 178,000*g* for 45 min at 4°C. The extract was passed through a 0.45-µm filter (Alltech, Deerfield, IL).

2.4.7 Tandem affinity purification

Purifications were performed as described by Rigaut *et al*. with some modifications. Briefly, total protein extract was incubated for 1 h at 4°C under gentle rotation with 500 µL IgG Sepharose 6 Fast Flow beads (GE-Healthcare, Little Chalfont, UK), pre-equilibrated with 10 mL extraction buffer. The IgG Sepharose beads were transferred to a 1 mL Mobicol column (MoBiTec, Göttingen, Germany) and washed with 10 mL IgG wash buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% NP-40, 5% ethylene glycol) and 10 mL tobacco (*Nicotiana tabacum* L.) etch virus (TEV) buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% [v/v] NP-40, 0.5 mM EDTA, 1 mM PMSF, 1 µM E64, 5% [v/v] ethylene glycol). Bound complexes were eluted via AcTEV digest (2x 100 U; Invitrogen) for 1 h at 16°C, followed by two wash steps with 750 µL calmodulin-binding buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% [v/v] NP-40, 10 mM β-mercaptoethanol, 1 mM imidazole, 2 mM $CaCl_2$, 1 mM magnesium acetate, complete EDTA-free protease inhibitor cocktail (Roche Diagnostics, Brussels, Belgium), 5% [v/v] ethylene glycol). The $CaCl_2$ concentration of the IgG-eluted fraction was adjusted to 2 mM, and incubated for 1 h at 4°C under gentle rotation with 500 µL calmodulin-agarose beads (Stratagene, La Jolla, CA), pre-equilibrated with 10 mL calmodulin-binding buffer. The calmodulin-agarose beads were packed in a Mobicol column and washed with 10 mL calmodulin-binding buffer. Bound complexes were eluted with 2.5 mL calmodulin elution buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% [v/v] NP-40, 10 mM β-mercaptoethanol, 1 mM imidazole, 25 mM EGTA, 5% [v/v] ethylene glycol), and precipitated with TCA (25% [v/v]). The protein pellet was washed twice with ice-cold aceton containing 50 mM HCl, redissolved in sample buffer and separated on 4-12% gradient NuPAGE gels (Invitrogen). Proteins were visualized with colloidal Coomassie Brilliant Blue staining.

2.4.8 Gel filtration

TAP-purified protein complexes, eluted from calmodulin-agarose beads, were separated at 4°C on a Superdex 200 (Omnifit 300/10; GE-Healthcare) size-exclusion column at a flow rate

of 75 µL/min, pre-equilibrated in gel filtration buffer (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 5 mM EGTA, 5 mM β-glycerophosphate, 1 mM NaF, 0.1 mM sodium vanadate, complete EDTA-free protease inhibitor cocktail) (Roche Diagnostics). Fractions of 500 µL were collected. Histone H1 kinase assay was carried out by incubating 25 µL of each fraction with 2 µCi [γP^{32}]ATP in the presence of 1 mg/mL histone H1 (Sigma-Aldrich), cAMP-dependent kinase inhibitor, 50 mM Tris-HCl (pH 7.8), 15 mM $MgCl_2$, 5 mM EGTA, and 1 mM DTT. After 20 min of incubation at 30°C, the reaction was stopped by heating the samples at 95°C in the presence of SDS-PAGE sample buffer. Samples were analyzed with 12% SDS-PAGE, stained with Coomassie Brilliant Blue, and autoradiographed. Next, 10 µL of the sized fractions were analyzed by immunoblotting with anti-CDKA;1 antiserum.

2.4.9 BN/SDS-PAGE

Protein solutions were subjected to buffer exchange at 4°C with the Ultra4 Amicon centrifugal device (Millipore, Bedford, MA) against BN sample buffer (30 mM HEPES, pH 7.0, 150 mM potassium acetate, 1% [v/v] protease inhibitor cocktail, 10% [v/v] glycerol). Protein concentration was determined by Bradford assays (Bio-Rad, Hercules, CA). For BN separation, 400 µg of protein was used from the total protein extract and the IgG-unbound fraction, and 55 µg from the TAP eluate. Digitonin (Calbiochem, La Jolla, CA) was added to a final concentration of 8.5 g/g protein. Samples were incubated on ice for 30 min, and insoluble material was removed by centrifugation at 20,000*g* for 45 min at 4°C. Native 5-16% gradient gels with a 4% stacking gel were cast on the SE600 Ruby system (GE-Healthcare). Running conditions were 45 min at 100 V/7 mA and 10-15 h at 500 V (max)/15 mA in blue cathode buffer (50 mM tricine, 15 mM BisTris, 0.2% Coomassie G-250, pH 7.0) and anode buffer (50 mM BisTris, pH 7.0) at 4°C. For separation in a second-dimension 12% SDS-PAGE, lanes from the first-dimension gel were cut out and incubated for at least 30 min in BN-denaturing buffer (50 mM Tris, pH 6.8, 66 mM Na_2CO_3, 10% [w/v] glycerol, 2% [w/v] SDS, and 2% β-mercaptoethanol). First-dimension strips were placed onto the stacking gel (4%) and overlaid with 0.5% agarose. The second-dimensions were run on the Ettan Dalt Six system (GE-healthcare) for 1 h at 600 V/400 mA/2.5 W, followed by 4.5 h at 16.6 W per gel. Immunoblotting of the second dimension was performed according to a standard protocol.

2.4.10 Proteolysis and peptide isolation

Complete lanes from the protein gels were cut into slices, collected in microtiter plates, and further processed for MS analysis as described before with minor modifications (Shevchenko et al., 1996). Dehydrated gel particles were rehydrated in 4 µL digest buffer containing 25 ng trypsin (MS Gold; Promega, Madison, WI), 100 mM NH_4HCO_3 and 10% (v/v) CH_3CN for 30 min at 4°C. After addition of 10 µL of a buffer containing 25 mM NH_4HCO_3 and 10% (v/v) CH_3CN, proteins were digested at 37°C for 3 h. The resulting peptides were concentrated and desalted with microcolumn solid phase tips (PerfectPureTM C18 tip, 200 nL bed volume; Eppendorf, Hamburg, Germany) and eluted directly onto a MALDI target plate (OptiMaldi; Applied Biosystems, Foster City, CA) with 1.1 µL of 50% CH_3CN:0.1% CF_3COOH solution saturated with α-cyano-4-hydroxycinnamic acid and spiked with 20 fmole/µL Glu1-fibrinopeptide B (Sigma-Aldrich).

2.4.11 Acquisition of mass spectra

A MALDI-tandem-MS instrument (4700 Proteomics Analyzer; Applied Biosystems) was used to acquire peptide mass fingerprints (PMFs) and subsequent 1 kV CID fragmentation spectra of selected peptides. PMFs and peptide sequence spectra were obtained with the settings presented in the Supplemental Tables 1 and 2. Each MALDI plate was calibrated according to the manufacturer's specifications. All PMF spectra were internally calibrated with the internal standard at *m/z* 1570.677 (fibrinopeptide B), resulting in an average mass accuracy of 5 ppm ± 10 ppm for each analyzed peptide spot on the analyzed OptiMALDI targets. Using the individual PMF spectra, up to eight peptides, exceeding a signal-to-noise ratio of 20 that passed through a mass exclusion filter (Supplemental Table 3) were submitted to fragmentation analysis. Fragmentation spectra were recorded according to the settings displayed in Supplemental Table 2 (no internal calibration was used).

2.4.12 MS-based protein homology identification

PMF and peptide sequence spectra of each sample were processed with the accompanied software suite (GPS Explorer 3.5; Applied Biosystems) with parameter settings as summarized in Supplemental Tables 4 and 5. Data search files were generated according to the settings presented in Supplemental Table 6 and submitted for protein homology identification by using a local database search engine (Mascot 2.1; Matrix Science, London,

UK). An in-house non-redundant database with acronym SNAPS (for Simple Non redundant Assembly of Protein Sequences; version 0.2) for *Arabidopsis* (72,161 sequence entries, 28,697,815 residues; available at http://www.ptools.ua.ac.be/snaps) was compiled from eight public databases (Supplemental Table 7). Protein homology identifications of the top hit (first rank) with a relative score exceeding 95% probability were retained. Additional positive identifications (second rank and more) were retained when the score exceeded the 98% probability threshold. The probability-based MOWSE score (Perkins et al., 1999), assuming that the observed match is significant (*P*-value = 0.05), had to equal 61 when submitting PMF data to the database, and 31 for individual peptide ions when submitting peptide sequence spectra. Peptide mass spectra of proteins that were solely identified on the basis of PMF or a single peptide sequence were annotated in the supplementary data set. Proteins belonging to a multiprotein family were singled out based on the identification of unique and diagnostic peptides. To estimate the false positive rate of the protein homology data set, a decoy database from the *Arabidopsis* SNAPS (version 0.2) was generated. Each protein amino acid sequence was shuffled with the EMBOSS shuffle tool (Rice et al., 2000). The score distribution and false positive frequency of the spectra from the different TAP experiments with the decoy database was plotted (Supplemental Figure 1).

2.4.13 Data analysis

Homology-based searches of the identified proteins were performed with the BLAST and PSI-BLAST algorithms. Gene Ontology (GO) annotation searches were done via the interface at The *Arabidopsis* Information Resource (TAIR) database (http://www.*Arabidopsis*.org). The Biological Networks Gene Ontology (BiNGO) (Maere et al., 2005) tool that is implemented as a plugin for Cytoscape (http://www.cytoscape.org), was used to determine which GO biological process was statistically overrepresented in the set of identified interactors, using the Hypergeometric test combined with a Benjamini and Hochberg false discovery rate correction (*P* cut-off at 0.05) (Maere et al., 2005). A protein-protein interaction network was built with the Cytoscape software.

2.5 Supplementary data

Supplemental material is available in the on-line version of this article at http://www.mcponline.org/cgi/content/full/M700078-MCP200/DC1 and is provided in the CD-ROM accompanying the thesis.

Acknowledgments

We thank Laszlo Bögre for providing the cell suspension culture; Michael Fromm for the vector ntapi.289.gw.gck; Eddy Esmans for support with MS; Mansour Karimi, Jan Zethof, Rebecca De Clercq, Wilson Ardiles, Karel Spruyt, Pierre Hilson, Jan Wuyts, Steven Maere, Stefaan Vandamme, Joanna Boruc, Jenny Russinova and the members of the "GBOU" users commission for experimental advice and assistance; and Martine De Cock for help in preparing the manuscript.

Notes to Chapter 2:

Chapter 3

Boosting tandem affinity purification of plant protein complexes

Contributions to Chapter 3 and the supplement linked to it:

The body of this chapter was entirely the work of Jelle Van Leene. The design, construction and evaluation of all discussed alternative TAP tags was to a large extend done by Jelle Van Leene, together with Guido Lopes dos Santos Santiago and Emilie De Witte as part of their training to obtain their masters degree in Biotechnology.

Chapter 3: Boosting tandem affinity purification of plant protein complexes

Jelle Van Leene, Erwin Witters, Dirk Inzé, and Geert De Jaeger

Abstract

Protein interaction mapping based on the tandem affinity purification (TAP) approach has been successfully established for several systems such as yeast and mammalian cells. As described in chapter 2, we managed to set up a technology platform based on the traditional TAP tag developed for yeast, enabling the isolation of protein complexes from plant cell suspension cultures and subsequent protein identification by mass spectrometry. Nevertheless, several limitations are still associated with the technology when using plant protein extracts, such as low complex yield and contamination of the TAP eluate by 'sticky' proteins. This is further reflected in the relatively few protein complex purifications reports in the plant research field. Here, we shortly review the TAP status in plants, highlight solutions for possible pitfalls and present a major breakthrough in the quest for a better TAP tag in plants.

A modified version of this chapter is published in Trends in Plant Science 13, 517-520 (2008)

3.1 The rise of TAP

Over the last 20 years, a wide variety of methods have been developed to explore protein interactions. Co-immunoprecipitation or yeast two-hybrid were often the method of choice, but the emergence of powerful, ultrasensitive high-throughput mass spectrometry (MS), together with the availability of comprehensive protein sequence repertoires, has favored the development of methods relying on *in situ* affinity purification of protein complexes. Especially, the tandem affinity purification approach based on the expression of a bait protein fused to a double affinity tag (the TAP tag), has proven to be of great value. The classical TAP tag consists of two immunoglobulin G (IgG)-binding domains of protein A from *Staphylococcus aureus*, a specific protease cleavage site for elution by addition of the tobacco (*Nicotiana tabacum*) etch virus (TEV) protease, and a calmodulin-binding peptide (CBP). Purification steps were optimized for highest recovery while maintaining protein complex integrity. TAP of protein complexes was first demonstrated in yeast (*Saccharomyces cerevisae*) (Rigaut et al., 1999) and was soon applied in a wide variety of organisms, giving rise to high-quality and comprehensive protein interaction networks (Gavin et al., 2006; Krogan et al., 2006). Nowadays, databases are filled with protein interaction data from TAP experiments, but in the plant research field, the TAP approach considerably lacks behind. Here, we review and discuss the use of TAP in plants and provide solutions to problems associated with the technology.

3.2 TAP in plants: an overview

Until now, only a limited number of purifications from plant material through TAP have been reported. Most have been performed in *Arabidopsis* (Rohila et al., 2004; Xing and Chen, 2006; Batelli et al., 2007; Van Aken et al., 2007; Van Leene et al., 2007; Dufresne et al., 2008; Takahashi et al., 2008; Zhao et al., 2008) and in rice (*Oryza sativa*) (Rohila et al., 2006; Abe et al., 2008) with the traditional yeast tag or with a plant-adapted version, called the improved TAP tag (TAPi) (Rohila et al., 2004). In this TAPi tag, cryptic splice sites and poly-A sites were removed and AT-or GC rich regions were adapted through third codon position changes. The duplicated protein A domain was made non-identical to reduce recombination and repeat-induced gene silencing. The castor bean catalase intron1 was introduced to increase gene expression both in monocots and dicots. This intron made it further possible to distinguish

gene expression in bacteria from that in plants. Transient expression in tobacco leaves of traditional TAP tag fusions to GFP revealed the existence of a context dependent nuclear localization signal (NLS) in the TAP tag, active only when the tag is fused to the N-terminus of GFP. ProtA-TEV-GFP and CBP-GFP fusions demonstrated that this NLS was present in the CBP. The CBP was subsequently mutated by replacement of basic amino acids with serine residues to remove the NLS and to keep the calmodulin-binding properties, giving rise to the TAPi tag. The authors also reported binding of an *Arabidopsis* protease to IgG beads, reducing the recovery of protease sensitive target proteins from the IgG-agarose during TEV cleavage. Analysis of different protease inhibitors with TAP fusions of β-glucuronidase, showed that the E-64 cysteine protease inhibitor increased recovery, and they proposed to add this inhibitor during the TEV cleavage, which is performed at 16°C, a temperature at which this protease is active. They further demonstrated that *in vivo* cross-linking with formaldehyde increases the recovery of less stable protein complexes. In *Oryza sativa*, protein interactions around 41 rice protein kinases were examined with this TAPi tag, expressed from the maize ubiquitin promoter (Rohila et al., 2006). After subtracting background proteins, 23 kinases showed interactions, leading to a success rate of 53 %. In our previous study (Chapter 2) (Van Leene et al., 2007), we developed a TAP-based technology platform to study the cell cycle interactome of *Arabidopsis*. In total 42 interactions were mapped with 6 core cell cycle baits. Instead of using plants, cell suspension cultures were chosen, as they are readily transformed and as they provide an unlimited supply of protein extracts from dividing cells. Purification of a protein complex from stably transformed *Arabidopsis* was first demonstrated with an alternative TAP-tag (TAPa) (Rubio et al., 2005). In this TAPa tag, the CBP domain was replaced with a 9xMyc and 6xHis sequence, making TAP complementary with purification of complexes that have cation-dependent interactions or activity, as there is no need for chelating agents in the final elution. It also allows an alternative during the second purification step, since one has the opportunity to use either the 9xMyc or the 6xHis tag, although a more efficient recovery was reported using IMAC, compared to myc epitope immunoprecipitation. Furthermore, the TEV cleavage site was replaced by the more specific and low-temperature active human rhinovirus 3C protease cleavage site, allowing all purification steps to be performed at low

temperature, thereby reducing the proteolytic activity in the extract. This TAPa-tag was fused to CSN3, enabling purification of the COP9 signalosome complex, although 6 eluates had to be pooled before MS analysis. Although this tag has often been used as an epitope tag for protein gel blotting (Kim et al., 2007) and in co-immunoprecipitation (Kim et al., 2007; Zentella et al., 2007) or chromatin immunoprecipitation experiments (Zentella et al., 2007), only a single protein complex has been characterized with the TAPa tag (Rubio et al., 2005). In conclusion, in plants, the traditional yeast TAP and the TAPi tags perform best so far.

3.3 TAP in plants: pitfalls and solutions

Approximately 10 years after the proof of concept in yeast, the few TAP data from plants demonstrate that problems are associated with the method. Indeed, unlike in yeast, efficient homologous recombination is not feasible in higher plants. So, the endogenous protein and the tagged counterpart will compete for complex assembly. To overcome this pitfall, different strategies can be followed. The TAP-tagged protein can be introduced into a mutant background, where the endogenous protein is suppressed by RNA interference (Forler et al., 2003) or is eliminated by T-DNA insertion (Rubio et al., 2005). These complementation approaches determine the functionality of the tagged protein and increase the success rate of the purification, because more interactors are available for complex assembly with the tagged protein. A more generic approach to increase competition is overexpression of the tagged bait, a strategy used in all successful TAP reports in plants so far. Another problem are false negative interactors, especially when low-abundant complexes are studied. As proteins are present in a high dynamic range, varying from only 10-100 copies to more than 10^7 copies per cell, and as they cannot be amplified like polynucleotides by PCR, the success rate of TAP depends on the amount of protein complexes purified and the MS sensitivity. One possibility to circumvent the problem of false negatives is the combination of multiple TAP eluates from parallel purifications (Rubio et al., 2005). Alternatively, the amount of protein extract can be increased before purification. When studying basic cell biological processes, plant cell suspension cultures have a major advantage compared to whole plants because they are fast growing and provide an unlimited supply of synchronizable biological material. Moreover, the PSB-D culture used previously (Van Leene et al., 2007), has a ploidy level of 8C, meaning that more proteins are

available for complex assembly. The suspension culture is ideal to investigate the cell cycle (Van Leene et al., 2007), but can also be used to isolate complexes involved in other fundamental processes such as primary metabolism, gene expression or cell wall synthesis.

3.4 The quest for a better TAP tag

Despite the valuable strategies described above, it was clear that a major leap forward would only be possible through further optimization of complex purification. Therefore, we evaluated different TAP tags for plant cells (Figure 1 and Chapter 3 supplement).

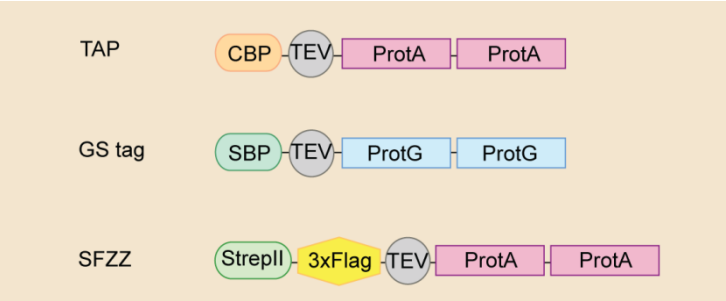


Figure 1. Overview of tested TAP tags: 3xFlag, three copies of Flag tag; ProtA, immunoglobulin G (IgG)-binding domain of protein A; ProtG, IgG-binding domain of protein G; SBP, streptavidin-binding peptide; StrepII, Strep-tag II.

In line with the TAPa tag (Rubio et al., 2005), we replaced the CBP part in the traditional TAP tag with linear peptide epitopes to reduce background, giving rise to the SFZZ tag (Figure 1). Production in transgenic cell suspension cultures of CKS1 as bait fused to the traditional TAP tag and the SFZZ tag was compared by western blot analysis (Figure 2). The anti-CKS1 antibody used in this experiment detected both tagged CKS1 bait and endogenous CKS1. Although the protein band intensities of the endogenous CKS1 indicate that a slightly higher amount of protein extract had been loaded for the SFZZ culture, we can conclude that protein accumulation levels were comparable (Figure 2).

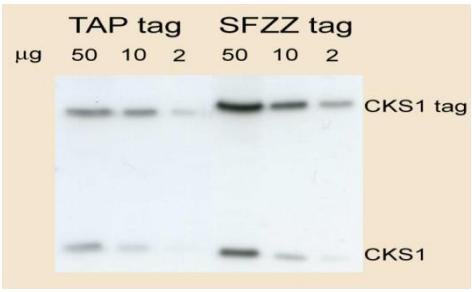


Figure 2. Western blot analysis of CKS1 production fused to the traditional TAP tag and the SFZZ tag in transgenic cell cultures: 50, 10 and 2 µg of total protein extracts of cultures expressing CKS1 tag were separated via SDS-PAGE. Both TAP and SFZZ tag fusions were analyzed via immunoblotting with antibodies against the bait protein. Endogenous protein levels are also shown.

During the first purification step, the protocol was identical to that used with the traditional tag. In the second step, anti-Flag resin was used instead of calmodulin, and bound complexes were released by competitive elution with triple Flag peptide. Bait recovery using the SFZZ tag was comparable to that obtained with the TAP tag, because the intensities of the protein bands representing the bait protein in the final eluate were similar, while an equal amount of bait was used as input (Figure 3).

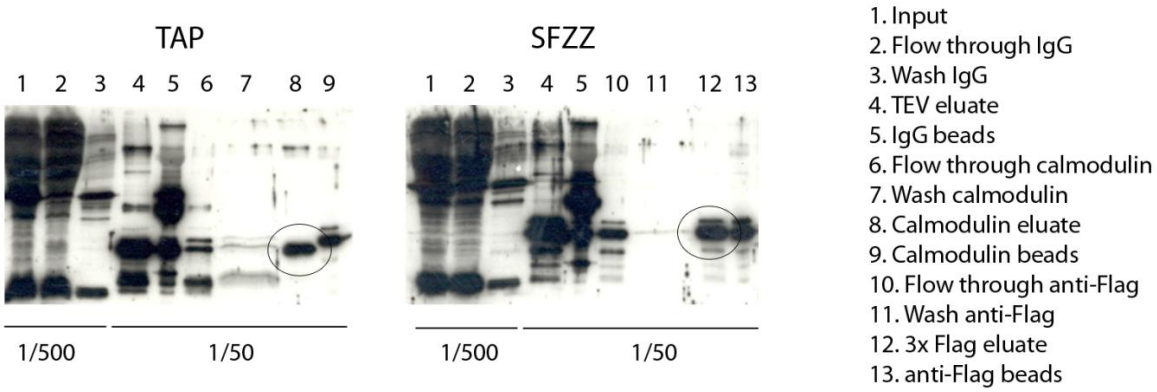
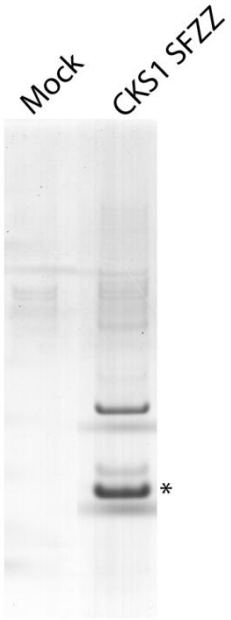


Figure 3. Western blot analysis following purification with the TAP tag and the SFZZ tag using CKS1 as bait. The fraction of each sample applied for SDS-PAGE is shown beneath the immunoblot. Protein bands corresponding to CKS1 present in the final eluates are encircled. Proteins were detected with anti-CKS. Human serum was added to prevent non-specific detection of the ZZ tag.



Although the final TAP eluates obtained with this SFZZ tag looked much 'cleaner' on gel as compared to the eluate with the traditional tag (Figure 4 and Chapter 2 Figure 5), only few interacting proteins could be sequenced (Table 1).

Figure 4. Evaluation of background and complex purification with CKS1 as bait using the SFZZ purification protocol. Final eluates were precipitated, separated on 4-12% NuPAGE gradient gels, and visualized with Coomassie G. The bait protein is indicated with an asterisk.

Table 1.

a) Confirmed proteins purified with CKS1 fused to SFZZ tag from 3 experiments.

Prey	*MW*	*Peptide count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT2G22490 CYCD2;1	41124	3	12	68/61	32/31
AT3G48750 CDKA;1	34123	17	56	712/61	88/31
AT4G10320 isoleucyl-tRNA synthetase	136368	15	14	119/61	37/31

b) Confirmed proteins purified with CKS1 fused to TAP tag from 3 purification experiments

Prey	*MW*	*Peptide Count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT3G53880 aldo/keto reductase	35185	9	38	79/61	
AT3G48750 CDKA;1	34123	7	31	312/61	92/31
AT3G54180 CDKB1;1	35524	4	20	178/61	99/31
AT1G20930 CDKB2;2	36052	2	8	79/61	49/31
AT1G47230 CYCA3;4	42647	4	14	110/61	78/31
AT4G14310 unknown	106008	9	12	219/61	60/31

Also the TAPa tag has probably to deal with this low complex yield, because 6 different TAP eluates had to be pooled to identify the COP9 signalosome complex (Rubio et al., 2005). These results indicate that replacement of the CBP domain in the traditional tag by linear epitope tags, leads to lower complex formation with the bait protein *in vivo* or lower stability of complexes containing bait protein during purification. Interference with proper complex assembly could be due to sticking of the basic His tag present in the TAPa tag to negative stretches on the bait protein or sticking of the acidic triple Flag tag repeat in the SFZZ tag to positive stretches on the bait protein. Similar problems observed with the SFHA tag, another evaluated alternative TAP tag, are discussed further in the supplement accompanying Chapter 3.

Despite a layer of background proteins sticking to the calmodulin resin, in our hands, the best results with respect to complex yield were, until recently, always obtained with the traditional or the TAPi tags (Rohila et al., 2004). Background proteins sticking to the resins and other false positives from non-specific binding to complexes after protein extract

preparation can be determined by mock and exogenous protein purifications, such as Green Fluorescent Protein (GFP) and β-glucuronidase. This list of proteins is then systematically subtracted from the original prey list. To get rid of bait-specific false positives absent in the control TAP list, it is also valuable to repeat purifications, and to give more confidence to interactions confirmed in multiple experiments (Van Leene et al., 2007), or give the best protein identification scores. Assigning confidence scores to interactions by integrating interaction data with other data sources is also rewarding, a method often applied in prediction of protein-protein interactions (Geisler-Lee et al., 2007; Cui et al., 2008).

3.5 A new TAP tag for plants: the GS tag

In our continuous search for an ideal TAP tag for plants, we evaluated the GS tag (Burckstummer et al., 2006) that combines two IgG-binding domains of protein G with a Streptavidin-binding peptide, separated by two TEV cleavage sites (Figure 1). This tag, developed to study mammalian protein complexes, has been reported to give a 10-fold increase in bait recovery compared to the traditional TAP tag. We adapted the GS protocol for plant cells and tested background levels by comparing two mock and two GFP purifications with the traditional TAP tag (Figure 5).

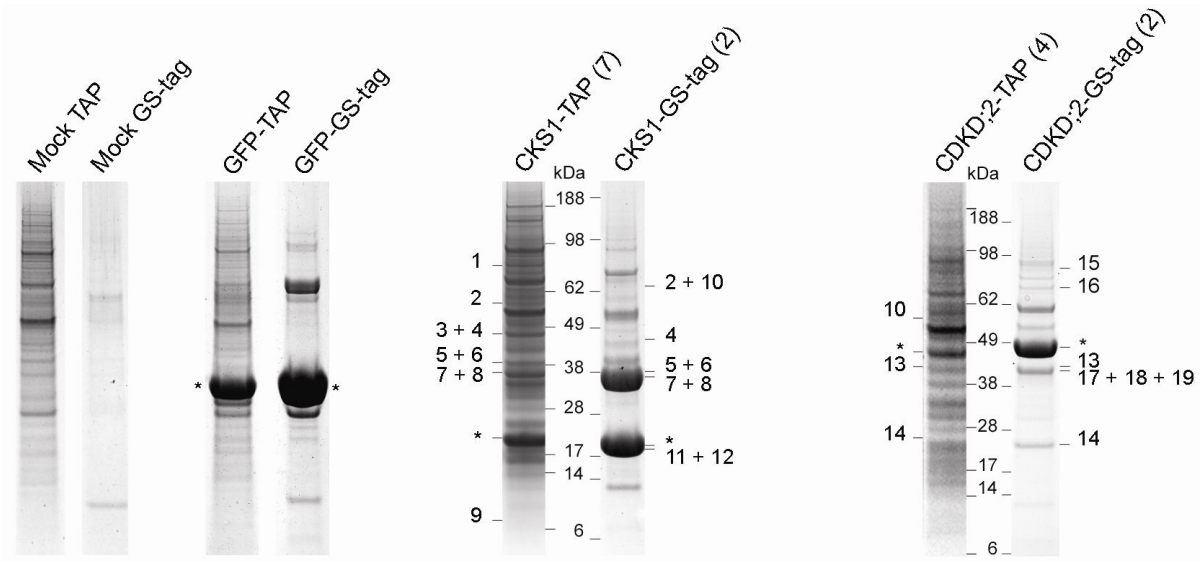


Figure 5. Comparison of final eluates with the traditional TAP and GS tag protocols: final eluates were precipitated, separated on 4-12% NuPAGE gradient gels, and visualized with Coomassie G. Background levels were analyzed with mock and GFP tag purifications. Proof of concept was demonstrated for CKS1 (At2g27960) and CDKD;2 (At1g66750). Preys identified via MALDI-TOFTOF and confirmed in multiple experiments are indicated (Table 2). The number of purifications used to determine the confirmed interactors is shown between

parentheses. *Asterisks* indicate bait proteins. 1) AT3G53880 aldo/keto reductase; 2) AT4G14310 unknown; 3) AT1G47230 CycA3;4; 4) AT2G22490 CycD2;1; 5) AT1G76540 CDKB2;1; 6) AT1G20930 CDKB2;2; 7) AT3G54180 CDKB1;1; 8) AT3G48750 CDKA;1; 9) Arath05g16630 (Eugene) unknown; 10) AT4G16143 importin alpha-2; 11) AT5G40460 unknown; 12) AT1G10690 unknown; 13) AT5G27620 CycH;1; 14) AT4G30820 MAT1; 15) AT1G03190 UVH6; 16) AT1G55750 TFIIH-related; 17) AT1G32380 PRS2; 18) AT2G35390 PRS1; 19) AT2G44530 PRS, putative.

Background levels, counted as the average number of proteins identified in two experimental repeats, dropped from 62 to 8 and from 87 to 11 proteins for mock and GFP purifications respectively, making MS analysis much less labour intensive and identification of genuine protein interactions easier, especially with low-abundant complexes.

An additional benefit of the GS tag is the higher cellular concentration levels of the bait protein (Figure 6) and the concomitantly higher complex incorporation and yield, as shown by the stronger A-type cyclin-dependent kinase (CDKA;1) band in the CDK subunit 1 (CKS1) GS-tag purification (Figure 5).

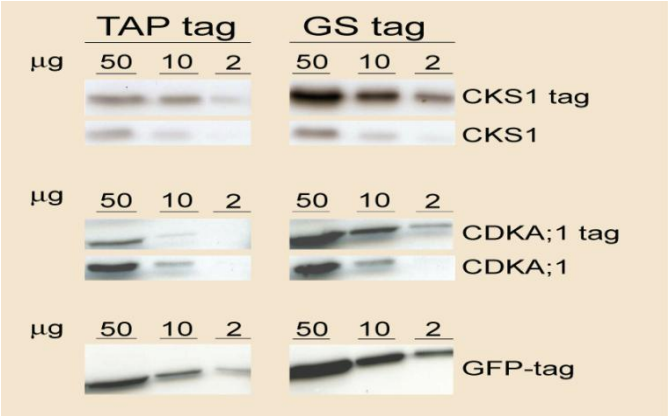


Figure 6. Higher bait expression with GS tag fusions: 50 µg, 10 µg, and 2 µg of total protein extracts of cultures expressing CKS1-tag, CDKA;1-tag or GFP-tag were separated via SDS-PAGE. Both TAP and GS tag fusions were analyzed via immunoblotting with antibodies against bait proteins. For CKS1 and CDKA;1, endogenous protein levels are also shown.

To further demonstrate the GS tag superiority, we present results obtained with two cell cycle baits, CKS1 and the D-type CDK-activating kinase CDKD;2 (Figure 5). Only the experimentally confirmed interactors are represented and compared with those obtained with the traditional TAP tag (Van Leene et al., 2007) (Table 2). For CKS1, most of the interactors confirmed previously in seven purifications with the traditional TAP tag, were found with the GS protocol with only two purifications. In addition, some new interesting interactions could be detected with the GS tag only. The known partners of CDKD;2, the H-type cyclin (CycH;1) and CDK-activating kinase assembly factor "ménage à trois" 1 (MAT1),

previously discovered in four purifications with the traditional TAP tag, were identified in two purifications with the GS protocol. Moreover, using the GS tag we demonstrate that as in rice (Rohila et al., 2006), the *Arabidopsis* CDKD;2 is part of the transcription factor IIH complex, because ultraviolet hypersensitive 6 (UVH6) and a TFIIH complex-related transcription factor co-purified. Furthermore, CDKD;2 might link regulation of cell division with nucleotide biosynthesis because of co-purification of three phosphoribosyl diphosphate synthetases (PRSs).

Table 2:

a) Confirmed proteins purified with CKS1 fused to GS tag from 2 purification experiments.

Prey	*MW*	*Peptide Count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT3G48750 CDKA;1	34160	16	49	585/61	86/31
AT3G54180 CDKB1;1	35524	5	24	277/61	115/31
AT1G76540 CDKB2;1	35741	10	33	252/61	80/31
AT1G20930 CDKB2;2	37223	7	23	235/61	80/31
AT2G22490 CYCD2;1	41124	15	49	551/61	99/31
AT4G14310 unknown	106008	11	15	292/61	80/31
AT4G16143 importin alpha-2	50097	11	31	72/61	
AT5G40460 unknown	12932	6	47	255/61	91/31
AT1G10690 unknown	12329	6	39	152/61	38/31

b) Confirmed proteins purified with CKS1 fused to TAP tag from 7 purification experiments.

Prey	*MW*	*Peptide Count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT3G53880 aldo/keto reductase	35185	9	38	79/61	
AT3G48750 CDKA;1	34123	7	31	312/61	92/31
AT3G54180 CDKB1;1	35524	4	20	178/61	99/31
AT1G76540 CDKB2;1	35741	14	38	83/61	
AT1G20930 CDKB2;2	36052	2	8	79/61	49/31
AT1G47230 CYCA3;4	42647	4	14	110/61	78/31
AT2G22490 CYCD2;1	41124	12	31	73/61	
AT4G14310 unknown	106008	9	12	219/61	60/31
Arath05g16630 Eugene unknown	3896	4	96	75/61	

c) Confirmed proteins purified with CDKD;2 fused to GS tag from 2 purification experiments.

Prey	*MW*	*Peptide Count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT5G27620 CYCH;1	38448	13	53	731/61	110/31
AT4G30820 MAT1	20149	12	47	426/61	112/31
AT2G35390 PRS 1	38647	7	26	108/61	43/31
AT1G32380 PRS 2	43819	7	22	137/61	43/31
AT2G44530 PRS, putative	42926	6	25	106/61	41/31
AT1G55750 TFIIH-related	67586	13	31	131/61	45/31
AT1G03190 UVH6	87093	31	44	1030/61	97/31

d) Confirmed proteins purified with CDKD;2 fused to TAP tag from 4 purification experiments.

Prey	*MW*	*Peptide Count*	*Seq coverage %*	*Protein score/ threshold*	*Best ion score/ threshold*
AT5G27620 CYCH;1	38448	9	33	361/61	89/31
AT4G16143 importin alpha-2	50097	2	20	59/61	47/31
AT4G30820 MAT1	20149	6	40	71/61	

Finally, we evaluated if purification with the GS tag recovered more bait protein compared to the traditional TAP tag when applying plant protein extracts, as it was reported in yeast (Burckstummer et al., 2006). For this, purifications were performed with GS and TAP tag fusions to GFP. First, expression levels of both fusions were determined by western blotting (data not shown), and based on this, an equal amount of bait protein and total protein was applied as input. To determine the bait recovery, fractions of the different purification steps were analyzed via western blotting (Figure 7). We did not observe a significant difference in final bait recovery (See EL2 Figure 7).

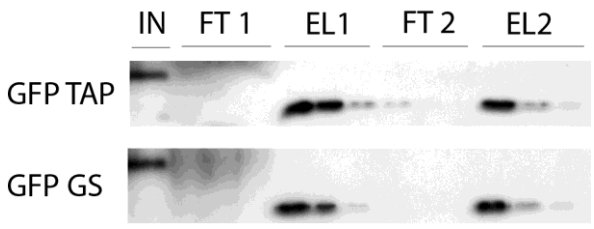


Figure 7: Similar bait recovery obtained with GS and TAP tag. For both purifications, an equal amount of bait protein was used as input as determined by western blotting. In addition, total soluble protein was adjusted to 50 mg in both purifications. Following fractions were analyzed via SDS-PAGE and western blotting with anti-GFP (1/1000) (+ 1/250 human serum

to block non-specific detection of tag): 10 µg of input and an equal fraction of the IgG flow through (FT1); TEV eluate (EL1), 2nd resin flow through (FT2), final eluate (EL2) 1/200, 1/500, 1/2000 of each fraction.

These results demonstrate that the improvements obtained with the GS tag are not due to higher bait recovery, but must be assigned to the increased bait protein accumulation levels and higher purification specificity reached with the GS tag. However, we can not exclude that the superiority of the GS tag is also caused by the physicochemical characteristics of the GS tag leading to less complex assembly interference compared to the TAP tag.

3.6 Conclusions and perspectives

We have shown that the GS tag outperforms the traditional TAP tag in plant cells, both concerning specificity and complex yield. Recently, we replaced the TEV protease cleavage sites in the GS tag with the rhinovirus 3C cleavage site for improved protein complex stability during purification. Combined with the latest and most sensitive MS technology, this tag should bring protein complex analysis in plants to its full bloom**.** Cloning with these tags is compatible with the Gateway system (Karimi et al., 2007) (Figure 8), and vectors for C- or N-terminal cloning are available at http://www.psb.ugent.be/gateway/.

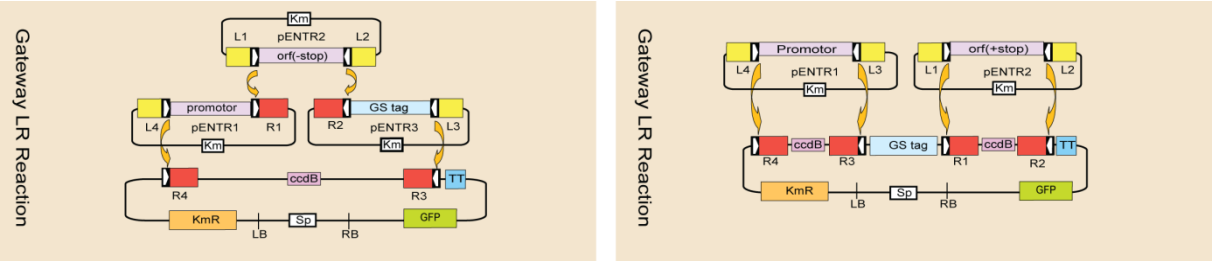


Figure 8 Schematic representation of the cloning strategy for GS fusions to the C-terminus or to the N-terminus of the bait protein. TT, cauliflower mosaic virus 35S transcription terminator; CcdB, toxic killer gene for negative selection; KmR, neomycin phosphotransferase II gene for selection of transformed plant cells; LB and RB, left and right border for T-DNA insertion.

3.7 Material and methods

3.7.1 Vector construction

The SFZZ tag was assembled by combining oligo's encoding the Strep-tag II, the triple Flag tag and the TEV cleavage site with a fragment containing the TEV cleavage site and the double Z-domain of Protein A amplified from the entry vector containing the TAP tag

(Chapter 2). The resulting fragment was cloned into pDonrP2RP3 by BP reaction (Gateway). The coding sequence of the GS tag was amplified by PCR from pCeMM-CTAP(SG) (Burckstummer et al., 2006) and cloned into pDonrP2RP3 by BP reaction (Gateway). Expression vectors were build by MultiSite LR reaction (Gateway) as described in Chapter 2. For the N-terminal GS destination vector, the NTAPi tag was removed from pKNTAP by *Pac*I/*Xho*I restriction digest and replaced with the coding sequence of the N-terminal version of the GS tag, amplified from pCeMM-NTAP(GS) (Burckstummer et al., 2006), together with the omega leader and Kozak sequence.

3.7.2 Cell culture transformation

Arabidopsis thaliana cell suspension cultures were transformed as described in Chapter 2.

3.7.3 Protein extract preparation

Plant material (15 g) of exponentially growing *Arabidopsis thaliana* cell cultures, ecotype *Landsberg erecta* (PSB-D), harvested two days after sub-culturing, was ground to homogeneity in liquid nitrogen. Crude protein extracts were prepared with 10 mL of extraction buffer (25 mM Tris-HCl, pH 7.6, 15 mM $MgCl_2$, 5 mM EGTA, 150 mM NaCl, 15 mM *p*-nitrophenylphosphate, 60 mM β-glycerophosphate, 0.1% (v/v) Nonidet P-40 (NP-40), 0.1 mM sodium vanadate, 1 mM NaF, 1 mM DTT, 1 mM PMSF, 10 µg/mL leupeptin, 10 µg/mL aprotinin, 5 µg/mL antipain, 5 µg/mL chymostatin, 5 µg/mL pepstatin, 10 µg/mL soybean trypsin inhibitor, 0.1 mM benzamidine, 1 µM *trans*-epoxysuccinyl-l-leucylamido-(4-guanidino)butane (E64), 5% (v/v) ethylene glycol) with an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC) at 4°C. The soluble protein fraction was obtained by a two-step centrifugation at 36,900*g* for 20 min and at 178,000*g* for 45 min at 4°C. The extract was passed through a 0.45-µm filter (Alltech, Deerfield, IL).

3.7.4 Tandem affinity purification with the SFZZ tag

Purifications were performed as described (Van Leene et al., 2007) with some modifications. Briefly, 200 mg of total protein extract was incubated for 1 h at 4°C under gentle rotation with 100 µL IgG Sepharose 6 Fast Flow beads (GE-Healthcare, Little Chalfont, UK), pre-equilibrated with 3 x 1 mL extraction buffer. The IgG Sepharose beads were transferred to a polyprep column (Biorad) and washed 3 times with 2 mL IgG wash buffer (10mM Tris-

HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 5% (v/v) ethylene glycol) and 1 mL tobacco (*Nicotiana tabacum* L.) etch virus (TEV) buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% (v/v) ethylene glycol). Bound complexes were eluted in an eppendorf in 100 µL TEV buffer via AcTEV (Invitrogen) digest (2x 100 Units, second boost after 30 min) for 1 h at 16°C. Eluate was collected by passing on a mobicol column and beads were washed two times with 150 µl Flag wash buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1mM benzamidine) and this wash step was collected together with the eluate. This eluate was incubated for 1 h at 4°C under gentle rotation with 400 µL anti-Flag M2 beads (Sigma-Aldrich), pre-equilibrated with 3 x 4 mL Flag wash buffer. Anti-Flag beads were transferred to a polyprep column (Biorad) and washed 3 times with 4 mL Flag wash buffer. Bound complexes were eluted in 2mL triple Flag elution buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1 µM E64, 1 mM PMSF, 1mM benzamidine, 100 µg/mL triple Flag peptide) (5 elution steps of 400 µL) and precipitated overnight on ice using TCA (25%v/v). The protein pellet was washed twice with ice-cold aceton containing 50 mM HCl, redissolved in sample buffer and separated on 4-12% gradient NuPAGE gels (Invitrogen). Proteins were visualized with colloidal Coomassie Brilliant Blue staining.

3.7.5 Tandem affinity purification with the GS tag

Purifications were performed as described (Van Leene et al., 2007) with some modifications. Briefly, 200 mg of total protein extract was incubated for 1 h at 4°C under gentle rotation with 100 µL IgG Sepharose 6 Fast Flow beads (GE-Healthcare, Little Chalfont, UK), pre-equilibrated with 3 x 1 mL extraction buffer. The IgG Sepharose beads were transferred to a polyprep column (Biorad) and washed with 10 mL IgG wash buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 5% (v/v) ethylene glycol) and 5 mL tobacco (*Nicotiana tabacum* L.) etch virus (TEV) buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% (v/v) ethylene glycol). Bound complexes were eluted in an eppendorf in 400 µL TEV buffer via AcTEV (Invitrogen) digest (2x 100 Units, second boost after 30 min) for 1 h at 16°C. Eluate was collected by passing on a mobicol column and beads were washed with 400 µl TEV buffer and this wash step was collected together with the eluate. This eluate was incubated for 1 h at 4°C under gentle rotation with 100 µL Streptavidin Sepharose High Performance beads (GE-Healthcare), pre-equilibrated with 3 x

1 mL TEV buffer. Streptavidin beads were transferred to a polyprep column (Biorad) and washed with 10 mL TEV buffer. Bound complexes were eluted in 1mL Streptavidin elution buffer (TEV buffer + 20 mM Desthiobiotin) (5 elution steps of 200 µL) and precipitated overnight on ice using TCA (25%v/v). The protein pellet was washed twice with ice-cold aceton containing 50 mM HCl, redissolved in sample buffer and separated on 4-12% gradient NuPAGE gels (Invitrogen). Proteins were visualized with colloidal Coomassie Brilliant Blue staining.

3.7.6 Mass spectrometry-based protein identification

Proteins were identified as previously described in Chapter 2.

Acknowledgments

We thank Giulio Superti-Furga and Tilmann Bürckstümmer for providing the plasmid encoding the GS tag, all members of the Functional Proteomics group at the Plant Systems Biology Department and all members of the Centre for Proteomics and Mass spectrometry at Antwerp University for technical assistance, and Martine De Cock for help in preparing the manuscript. This work was supported by grants from the Institute for the Promotion of Innovation through Science and Technology in Flanders (Generisch Basisonderzoek aan de Universiteiten [GBOU] grant no. 20193) and from the European Commission (Agronomics; LSHG-CT-2006-037704).

Chapter 3 supplement: Evaluation of alternative tandem affinity purification tags

3.8 Introduction

As discussed in chapter 3, protein complex purification from plant cells with the traditional TAP tag suffers from high background levels and low protein complex yield. So, further optimization of complex purification was required to bring protein complex analysis in plants to its full potential. This goal was reached by the introduction of the GS tag. However, preceding to this success, we tested a whole range of different TAP tags, such as the previously mentioned SFZZ tag (Chapter 3). This supplement gives an overview of all other evaluated TAP tags. Obtained results are discussed and provide further insight into the compatibility of affinity tags with purification of protein complexes from plant protein extracts. Promising results were booked with the CSFH tag, which could serve as an alternative where the GS tag fails. At the end of this supplement, future perspectives for further optimization are discussed based upon the obtained knowledge.

3.9 Results

For the design of new TAP tags, we focused on small epitope tags for which an affinity resin was commercially available (Hearn and Acosta, 2001; Terpe, 2003). Because of their small size, one would expect minimal interference with folding, localization, complex assembly and function of the target protein. First of all, we conducted a sequence homology-analysis, evaluating the uniqueness of these commercially available tags against the *Arabidopsis thaliana* proteome, because the most unique ones are thought to generate the lowest background. This analysis revealed that both the Strep-tag II and the Hemagglutinin (HA) tag were the most unique. In one TAP tag, we combined both tags with the TEV cleavage site. Also the Flag tag was implemented, as good results were obtained using a combination of this Flag tag with the double Z-domain of protein A in mammalian cells (Knuesel et al., 2003). So initially, two alternative TAP tags were designed and evaluated. Results obtained with one of these, namely the SFZZ tag, were described earlier (Chapter 3) (Van Leene et al., 2008). The other one, the SFHA tag, combines the Strep-tag II with 3 tandem repeats of the Flag tag and 3 tandem repeats of the HA tag, separated by two copies of the TEV protease cleavage site (Figure 1). For the epitope tags we chose to add 3 tandem repeats to maximize trapping of the bait on the affinity resin by avidity.

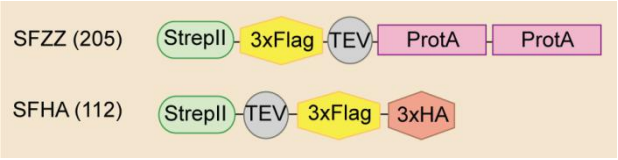


Figure 1: Schematic representation of the first two alternative TAP tags. The size of the tag is shown between parentheses as the number of amino acids.

With the SFHA tag, immuno-affinity resins can be used during the first purification step, while the Strep-tag II allows purification using Strep-Tactin beads. Due to the monoclonal character of the anti-HA or anti-Flag resins, one could expect reduction of non-specific binding. Competitive elution using triple Flag or triple HA peptide decreases purification time, however the TEV cleavage site was still included, although now a double repeat was introduced, to obtain a more efficient cleavage.

To evaluate the performance of the SFHA tag, different purification strategies were tested. Although binding to anti-HA and anti-Flag was satisfying, peptide elution from the corresponding resins was inefficient, leaving TEV cleavage as the only efficient elution

possibility. Consequently, purification through the Strep-tag II had to be used in the second purification step. However, we observed weak binding to Strep-Tactin, although elution using an excess desthiobiotin was efficient. Because anti-Flag M2 is cheaper than anti-HA, we decided to perform a tandem affinity purification with anti-Flag M2 in the first step, followed by TEV elution, binding to Strep-Tactin and elution with desthiobiotin. In general, purification with the SFHA tag was more specific due to the avoidance of calmodulin in the second purification step. However, complex recovery yield was even lower compared to the traditional TAP tag, most likely due to loss of complexes during Strep-Tactin binding. This was further supported by the absence of cell cycle-related or other interactors in the final eluates as identified by MS (data not shown). Another possible explanation for the failure of the SFHA and SFZZ tag (Chapter 3) was recently obtained by analyzing the primary sequence of the two tags. This learned that the Flag tag contains multiple Aspartate residues. By combining the StrepII tag with 3 Flag tags, a very negatively charged peptide stretch was generated (pI 3.5). In the SFHA and the SFZZ tag, this triple repeat is close to the C-terminus of the bait protein. Moreover, both tags are separated by a $Glycine_4Serine$-linker, making them very flexible. These tags might thus be electrostatically attracted by positive stretches on the bait protein, which is certainly the case for CKS1 (pI 9.7). Consequently, sticking of the tag to the bait might interfere with correct complex assembly, explaining the unsatisfying results obtained with SFHA and SFZZ tags.

Because our first two alternative TAP tags did not perform better compared to the traditional tag, we designed and tested 3 additional tags (Figure 2).

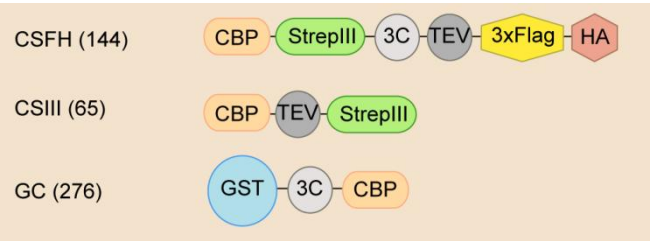


Figure 2: Schematic representation of the second set of alternative tags. The size of the tag is shown between parentheses as the number of amino acids.

The CSFH tag combines properties of the traditional TAP tag with the SFHA tag. We retained the triple Flag for the first purification step, because concerning background and bait purification yield, it performed well in the SFHA tag. Besides the TEV cleavage site, we

introduced the highly specific human Rhinovirus 3C cleavage site, allowing cleavage at 4°C. Carboxy-terminal of the triple Flag tag, we retained one copy of the HA tag as this provides a sensitive and specific detection tool of the bait (Figure 2). In the second purification step, either the CBP tag can be used, or an optimized version of the Strep-tag II with improved binding capacity for Strep-Tactin, namely the Strep-tag III (Junttila et al., 2005). Different purification strategies were tested, but the best results concerning bait recovery were obtained with anti-Flag M2 in the first step, and Strep-Tactin in the second step, followed by elution with an excess desthiobiotin. Between both purification steps, the most optimal elution of bound complexes was obtained by a combination of Rhinovirus 3C protease cleavage and elution with Flag peptide (Figure 3). We estimated a final bait recovery of 10 % by comparing the input protein band (CKS1-CSFH in lane 1) with that one in the final eluate (CKS1-CS in lane 9) (Figure 3), which is twice the yield as reached with the traditional TAP tag (data not shown).

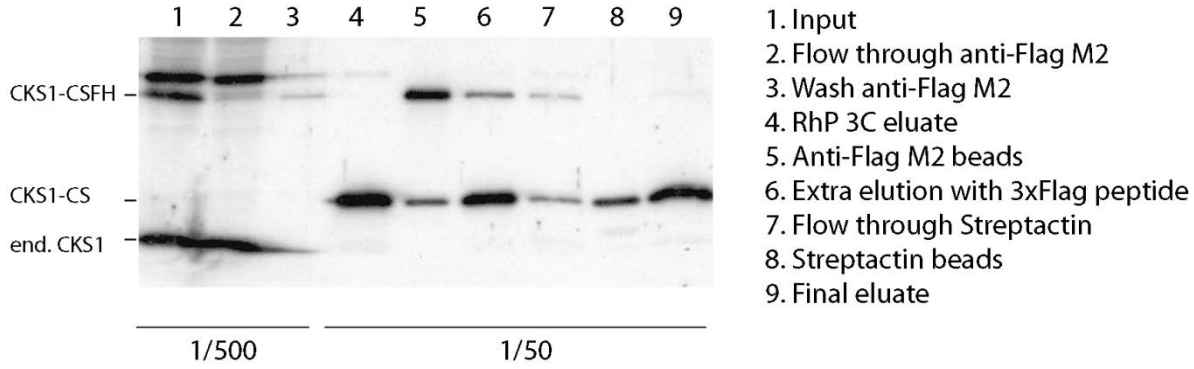


Figure 3. Western blot analysis with anti-CKS following purification with the CSFH tag using CKS1 as bait. In the first step, anti-Flag M2 resin was used, followed by elution with 3C protease, and purification on Strep-Tactin beads. This demonstrates efficient binding in both purification steps. Additional washing of the anti-Flag beads with buffer containing triple Flag peptide, increased recovery. The fraction of each sample applied for SDS-PAGE is shown beneath the immunoblot. CKS1-CS refers to the fusion protein after protease cleavage, end. CKS1 to the endogenous CKS1 protein.

The optimized purification protocol was applied to isolate complexes around two cell cycle baits, CKS1 and CycD3;1 and background was determined by mock purification on non-transformed cell culture (Figure 4).

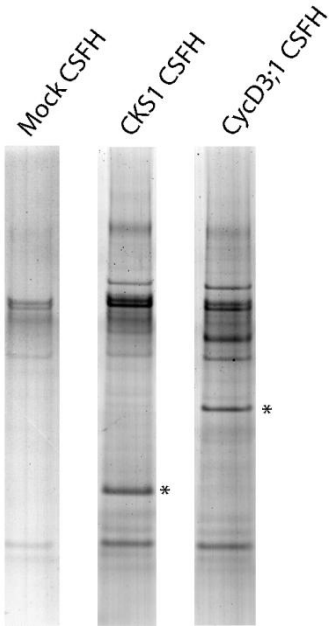


Figure 4. Evaluation of background and complex purification with CKS1 or CycD3;1 as bait using the CSFH purification protocol. Final eluates were precipitated, separated on 4-12% NuPAGE gradient gels, and visualized with Coomassie G. Bait proteins are indicated with an asterisk.

Interactors identified via MS were compared to those found with the traditional TAP tag (Chapter 2 Table 1). For CKS1 the results were not satisfying, i.e. only one CDK and no cyclins could be found as co-purified proteins (Table 1), while several B-type CDKs and cyclins were found with the traditional tag. This could be explained by the low protein accumulation levels obtained with the CSFH tag fused to CKS1 compared to the levels obtained with the TAP fusion (data not shown). However with CycD3;1, we demonstrate that the CSFH tag allows efficient complex isolation, as it co-purified the cell cycle proteins CDKA;1, CKS2, KRP6, SMR4, SMR6 and other interesting proteins, like AT4G02110 (Table 2). With the traditional tag on the contrary, only CDKA;1, CKS2, SMR4 and KRP6 were found here (Chapter 2 Table 1). AT4G02110 is similar to yeast Dpb11 or human TopBP1, which is not only an essential protein for the initiation of replication, but in addition a central dynamic adaptor for the cell cycle checkpoint response and DNA repair (Schmidt et al., 2008). Furthermore, it contains 3 CDK consensus phosphorylation sites, making it a good candidate substrate of CDKA;1/CycD3;1 complexes.

Like in the SFHA and SFZZ tags, the negatively charged triple Flag repeat is also present in the CSFH tag. However, here it is separated from the bait by 109 amino acids. This distance could explain why we found no interference with complex assembly with the CSFH tag fused to CycD3;1. With CKS1 however, this interference could still exist due to its extreme high pI, providing an additional explanation for the somewhat negative results compared to the TAP tag with CKS1.

Table 1: List of proteins co-purified with CKS;1 using the alternative CSFH tag.

Locus	*Protein name*	*Peptide Count*	*Protein Score*	*Best Ion Score*
AT3G48750	CDKA;1	5	97	37
AT2G27960	CKS1	6	174	68
AT5G40920	disease resistance protein-like	17	66	
AT1G26580	expressed protein similar to putative MYB TF	7	70	
AT2G31920	expressed protein	1	29	29
AT1G59820	haloacid dehalogenase-like hydrolase protein	14	63	
AT2G37370	hypothetical protein	15	76	
AT3G58160	myosin heavy chain, putative	18	70	

Table 2: List of proteins co-purified with CycD3;1 using the alternative CSFH tag.

Locus	*Protein name*	*Peptide Count*	*Protein Score*	*Best Ion Score*
AT4G02110	BRCT domain-containing protein	18	69	
AT3G48750	CDKA;1	23	722	72
AT2G27970	CKS2	7	228	55
AT1G07910	expressed protein	9	65	
AT1G36310	expressed protein	7	65	
AT3G19150	KRP6	3	53	36
Arath05g28620	plant transposase	6	63	
AT4G19110	protein kinase, putative	5	61	45
AT5G50180	protein kinase, similar to protein kinase ATN1	4	73	45
AT5G02220	SMR4	2	62	48
AT5G40460	SMR6	2	69	32
AT5G54770	thiazole biosynthetic enzyme, (ARA6) (THI1)	6	73	41
AT1G52450	ubiquitin carboxyl-terminal hydrolase-related	20	72	
Arath03g35880	unknown	7	62	
AT2G47090	zinc finger (C2H2 type) family protein	11	64	

Next, the Strep-tag III was implemented in the CSIII tag (Figure 2), together with two copies of the TEV cleavage site and the CBP tag. After optimization of the purification protocol, we observed a lot of non-specific proteins in the final eluate (Figure 5), confirming that the use of calmodulin generates a high false positive rate.

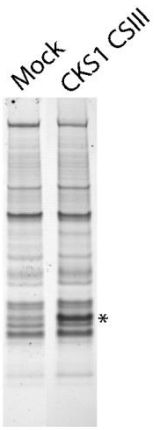


Figure 5. Evaluation of background and complex purification with CKS1 as bait using the CSIII purification protocol. Final eluates were precipitated, separated on 4-12% NuPAGE gradient gels, and visualized with Coomassie G. Bait proteins are indicated with an asterisk.

Finally, we developed the GC tag consisting of the CBP tag and the GST tag, separated by the human Rhinovirus 3C protease cleavage site (Figure 2). As no protein expression could be detected with 3 GC fusions (CKS1, CycD3;1, CDKD;2), we speculate that this tag does not fold properly leading to proteolytic degradation, and no further experiments were performed with this tag.

3.10 Conclusions and future perspectives

Different alternative tandem affinity purification tags were evaluated, as summarized in Table 3.

Table 3: Overview of tested dual affinity tags. Specific characteristics are shown in the last column. N/A = not assessed.

TAG	Modules	Expression	Bait recovery	Background	Complex retrieval	Other
TAP	CBP-TEV-ProtA2	+	+	high	good	TEV cleavage at 16°C
GS	SBP-TEV-ProtG2	++	+	low	good	TEV cleavage at 16°C
SFZZ	StrepII-3xFlag-TEV-ProtA2	++	+	low	bad	too low pI
SFHA	StrepII-TEV-3xFlag-3xHA	++	+	low	bad	small tag, specific detection
CSFH	CBP-StrepIII-3C-TEV-3xFlag-HA	+	++	low	good	cleavage at 4°C
CSIII	CBP-StrepIII	+	+	high	bad	no cleavage required
GC	GST-CBP	-	N/A	N/A	N/A	

With one of them, the CSFH tag, promising results were obtained, offering an alternative when the GS tag fails. Although results from the other alternative TAP tags were often not satisfying (SFZZ, SFHA, CSIII, GC), their analysis provided us with a lot of insight. For example, calmodulin generates a lot of background in plants, triple Flag repeats interfere with proper complex assembly due to their low pI when they are close to the bait protein. The SBP peptide of the GS tag on the other hand, has a normal pI of ± 6.0 and moreover, a computational analysis revealed the presence of an α-helical stretch, giving it secondary structure. These features might decrease interference with complex assembly. The same holds true for the original TAP tag. Although speculative, one could take into account these observations when designing dual-affinity tags.

As discussed in Chapter 3, we optimized protein complex isolation from plant cells by introducing the GS tag (Van Leene et al., 2008). Although false positives are limited with the GS tag, false negatives still remain a problem. For example, using the GS protocol with CKS1 as bait (Chapter 3 Table 2), we identified different B-type CDKs but no B-type cyclins, probably due to the unstable character of these cyclins. However, when B-type cyclins were used as bait, we did identify CKS1 and B-type CDKs. Currently, we are evaluating an optimized version of this GS tag, where the TEV cleavage site is replaced by the more specific and low-temperature active Rhinovirus 3C cleavage site (Figure 6; GS3C). This tag should decrease protein degradation during purification, as most plant proteases are not active at 4°C, leading to improved identification of unstable proteins.

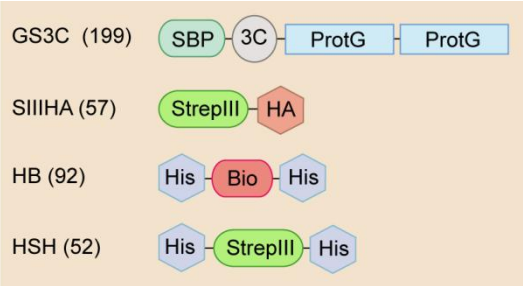


Figure 6: Schematic representation of TAP tags currently under evaluation. The size of the tag is shown between parentheses as the number of amino acids.

Finally, weak and transient interactors add another level of complexity: these kind of interactors dissociate often during extract preparation, when protein complexes are diluted and binding constants change, or during the long purification method. This may be partially prevented by shortening the purification protocol. For this purpose, we are evaluating the

SIIIHA tag (Figure 6), a short derivative of the CSFH tag. This tag is very small compared to the GS tag and does not require a time-consuming protease cleavage step. Moreover, it contains only one copy of the HA tag, anticipating on the inefficient peptide elution observed with the SFHA tag containing three tandem repeats of the HA tag.

Another way to tackle the problem of weak and short interactors is *in vivo* cross-linking of proteins prior to lysis. During this process, neighboring proteins are covalently attached to each other. By using formaldehyde or other cross-linkers with short spacers, non-specific cross-linking is limited. An additional advantage of formaldehyde is that it easily permeates into the plant cell. Once protein complexes are covalently stabilized, they can be purified under stringent conditions, decreasing false positives. A major problem however associated with the technique is that the TAP tag may lose its functionality as it gets covalently attached to other proteins or formaldehyde is changing the structure of the binding surface of the tag that interacts with the affinity resin. The HB tag (Figure 6), consisting out of a naturally occurring peptide that gets biotinylated (Bio) *in vivo*, flanked at both ends by the hexahistidine tag, has proven its compatibility with cross-linking in yeast (Guerrero et al., 2006; Tagwerker et al., 2006). We are currently investigating its usefulness in plant cells. This tag could also be used in chromatin-immunoprecipitation experiments to unravel transcription regulatory networks, as these experiments require cross-linking of proteins to DNA. In line with the HB tag, we are evaluating the HSH tag in which the Bio tag is replaced by the Strep-tag III (Figure 6). Because this tag does not contain any arginine or lysine residues, essential amino acids for cross-linking (Juergen Cast, personal communication), it may provide an equivalent alternative when the HB tag fails.

3.11 Materials and methods

3.11.1 Vector construction

The different TAP tags were assembled by combining oligo's encoding tags with fragments amplified by PCR. The resulting assemblies were cloned into pDonrP2RP3 by BP reaction (Gateway). Expression vectors were build by MultiSite LR reaction (Gateway) as described in Chapter 2.

3.11.2 Cell culture transformation

Arabidopsis thaliana cell suspension cultures were transformed as described in Chapter 2.

3.11.3 Tandem affinity purification with the CSFH tag

Extract preparation and purifications were performed as described (Van Leene et al., 2007) with some modifications. Briefly, extract was prepared in Flag extraction buffer. This is the same extraction buffer as used with the traditional TAP tag or with the GS tag, however without the reducing agent DTT to avoid reduction of S-bridges of the anti-Flag M2 antibodies. Next, 200 mg of total protein extract was incubated for 1 h at 4°C under gentle rotation with 400 µL anti-Flag M2 beads (Sigma-Aldrich), pre-equilibrated with 3 x 4 mL Flag extraction buffer. The anti-Flag beads were transferred to a polyprep column (Biorad) and washed 3 times with 4 mL Flag wash buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1mM benzamidine) and one time with 4 mL Rhinovirus protease buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF). Bound complexes were eluted in an eppendorf in 800 µL Rhinovirus protease buffer via a Prescission protease (GE-Healthcare) digest (2x 100 Units, second boost after 30 min) for 1 h at 4°C. Eluate was collected by passing on a mobicol column and beads were washed with 3 times with 800 µl Flag peptide elution buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1 µM E64, 1 mM PMSF, 1mM benzamidine, 100 µg/mL triple Flag peptide) and this wash step was collected together with the eluate. This eluate was incubated for 1 h at 4°C under gentle rotation with 800 µL Strep-Tactin beads (IBA, Göttingen), pre-equilibrated with 3 x 8 mL Strep Wash buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 2.5 mM EDTA, 2 mM DTT, 1 µM E64, 1 mM PMSF). Strep-Tactin beads were transferred to a polyprep column (Biorad) and washed 3 times with 8 mL TEV buffer. Bound complexes were eluted in 2mL Strep-Tactin elution buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 0.1% NP-40, 2 mM DTT, 1 µM E64, 1 mM PMSF 20 mM Desthiobiotin) (5 elution steps of 400 µL) and precipitated overnight on ice using TCA (25%v/v). The protein pellet was washed twice with ice-cold aceton containing 50 mM HCl, redissolved in sample buffer and separated on 4-12% gradient NuPAGE gels (Invitrogen). Proteins were visualized with colloidal Coomassie Brilliant Blue staining.

3.11.4 Mass spectrometry-based protein identification

Proteins were identified as previously described in Chapter 2.

Acknowledgments

We are grateful to Guido Lopes dos Santos Santiago, Emilie De Witte, to all members of the functional proteomics group at the Department of Plant Systems Biology and to all members of the Centre for Proteomics and Mass spectrometry at Antwerp University for technical assistance. We also would like to thank Savvas Savvides for analysis of the amino acid sequences of the tags. This work was supported by grants from the Institute for the Promotion of Innovation through Science and Technology in Flanders (Generisch Basisonderzoek aan de Universiteiten [GBOU] grant no. 20193).

Notes to Chapter 3 & supplement:

Chapter 4

Identification of the cell cycle interactome of *Arabidopsis thaliana* through a targeted proteomics approach

Contributions to Chapter 4

The experimental results presented in this chapter are the result of an extensive collaboration between the functional proteomics group of the Department of Plant Systems Biology lead by Geert De Jaeger, and the Centre for Proteomics and Mass Spectrometry of the University of Antwerp (Ceproma) lead by Erwin Witters. The former group was responsible for all cloning, transformation and purification steps, while protein identifications were performed by Dominique Eeckhout at Ceproma. For data analysis, Jelle Van Leene collaborated with Jens Hollunder, Stefanie De Bodt and Steven Maere from the Bio-informatics group of the Department of Plant Systems Biology. The body of this chapter was entirely written by Jelle Van Leene.

Chapter 4 preface: A brief introduction to the cell cycle of *Arabidopsis thaliana*.

Cell division is the process during which DNA and other cellular contents are duplicated (Interphase) and divided over two new daughter cells (Mitosis). The cycle of duplication and division is known as the cell cycle. The interphase is divided into the S-phase, during which cells replicate their DNA and two gap phases (G1 and G2), separating the S-phase from the M-phase. At the end of both gap phases a cell decides whether or not the transition to the next phase will take place. These decisions depend on intrinsic and extrinsic cues for instance when growth conditions are unfavourable due to a lack of energy or when the DNA is too much damaged and extra time for DNA repair is required, transition to the next phase is blocked. So the control of these transition points demands an extremely well regulation. During M-phase the nuclear envelope breaks down allowing separation of the duplicated DNA in mitosis. Finally, at the end of M-phase, two new cells are formed by a process termed cytokinesis.

Iteration of this cell cycle followed by cell growth and differentiation enables the existence of multicellular organisms like plants. In contrast to animals, plant development is largely post-embryonic (Inze and De Veylder, 2006). Cell divisions occur at specialized zones known as meristems, which are small niches of pluripotent cells. Leaves and flowers are derived from the shoot and floral meristems respectively, while (lateral) roots arise from the (lateral) root meristem and vascular tissues originate from the vascular cambium. These meristems are a source of new cells that eventually differentiate forming specific organs. However, many differentiated plant cells can dedifferentiate and regain their pluripotentiality, a feature that provides huge plasticity to plant development, necessary as plants have to deal with a lot of biotic and abiotic stress factors because of their sessile lifestyle. Another specific feature of plants is the presence of rigid cell walls preventing cell migration. Because plant cells do not move and are surrounded by this cell wall, cell division rates and patterns – i.e. determination of the cell division plane - are directly responsible for generating new

structures throughout development (Di Laurenzio et al., 1996). In addition, because of this cell wall, plants evolved unique mechanisms to split a cell in two. This process involves specific cytoskeletal arrays named the preprophase band and the phragmoplast. The complexity and flexibility of plant development is probably reflected in the huge number of molecules regulating cell division, and a characterization of the molecular protein complexes involved in cell division will offer valuable information to get a better view on plant growth and development.

The basic underlying mechanisms that govern the plant cell cycle are conserved among all eukaryotes and involve Cyclin-dependent kinase (CDK)/cyclin (CYC) complexes. Activation of these complexes allows rapid initiation of waves of phosphorylation events on a plethora of downstream substrates required for proper cell division. They are key players at G1/S and G2/M restriction points triggering onset of DNA replication and mitosis, respectively. The CDKs are the catalytic subunits that phosphorylate a consensus motif (a Serine or Threonine residue followed by a Proline) on the target substrate (Inze and De Veylder, 2006), while specificity of these complexes is determined by the regulatory cyclin partner.

Cyclin-dependent kinases

Arabidopsis thaliana possesses different subclasses of CDKs discriminated by their cyclin-binding domain. Based on the presence of the archeal PSTAIRE hallmark in its cyclin-binding motif, CDKA;1 is considered as the ortholog of yeast Cdc28/Cdc2 and human CDK1/2. This constantly expressed protein (Magyar et al., 1997) is a key regulator of both transitions points (Hemerly et al., 1995; Porceddu et al., 2001). As no orthologs of the human G1/S CDK4/6 were found in *Arabidopsis*, one could assume that CDKA;1 is the G1/S CDK in *Arabidopsis*.

A second class of CDKs, the B-type CDKs, are characterized by the presence of a PPTALRE or PPTTLRE signature in their cyclin-binding motif, reflecting the existence of two CDKB1 and two CDKB2 proteins (Vandepoele et al., 2002). In contrast to CDKA;1, expression of the 4 B-type CDKs is cell cycle-regulated with CDKB1 expressed at S, G2 and M-phase, while CDKB2 transcripts accumulate mainly at the G2/M boundary(Menges et al., 2005). The protein accumulation levels nicely follow the transcript levels, and a peak of kinase activity reaches a

maximum during mitosis. Other experiments confirmed the requirement of CDKB activity to progress into mitosis (Porceddu et al., 2001; Boudolf et al., 2004a), and in addition, CDKB1 also controls cell cycle exit (Boudolf et al., 2004b).

For full activity, CDKs require phosphorylation of a Threonine residue inducing a conformational change necessary for recognition of substrates. This phosphorylation is performed by CDK-activating kinases (CAKs). In *Arabidopsis* there are 4 CAKs, subdivided into CDKDs, a family of 3 CDKs related to vertebrate CAKs, and 1 plant-specific CDKF (Vandepoele et al., 2002). The CDKDs are functionally differentiated from CDKF by their cyclin dependence and substrate specificity: while CDKDs require an H-type cyclin (CYCH;1), CDKF;1 can phosphorylate its substrates without cyclin binding. Furthermore, CDKDs are able to phosphorylate both CDKA;1 and the C-terminal domain (CTD) of the largest subunit of RNA polymerase II, while CDKF;1 has been shown to activate CDKA;1 and CDKDs (Shimotohno et al., 2004; Umeda et al., 2005; Shimotohno et al., 2006). Because of the latter activation, CDKF;1 is also considered as a CAK-activating kinase (CAKAK). Through their regulation of CDK activity, these CAKs integrate hormonal signals into the cell cycle and play a role in determination of growth rate and differentiation status (Yamaguchi et al., 2003).

Next, plants posses 2 CDKCs and 1 CDKE. The CDKCs interact with T-type cyclins and transcripts are found mainly in differentiated tissues (Inze and De Veylder, 2006). Presumably, they play a role in transcription elongation by phosphorylation of the CTD of RNA polymerase II and in differentiation (Fulop et al., 2005). Furthermore, CDKC;2 is important in coupling transcription to splicing (Kitsios et al., 2008). CDKE;1 or HUA ENHANCER3 also phosphorylates the CTD of RNA polymerase II, however, in contrast to CDKCs, it is produced in dividing cells, and it is believed to act in cell expansion in leaves and cell-fate specification in floral organs (Wang and Chen, 2004). Finally, additional sequence analysis discovered 2 CDKGs, homologous to the human galactosyltransferase-associated protein kinase p58/GTA, and 15 new CDK-like genes that require further functional annotation.

Cyclins

As mentioned before, cyclins provide substrate specificity to the catalytic CKDs. Originally, they were named like this because their transcript and protein accumulation levels fluctuate during cell cycle progression. Unexpectedly, plants contain more cyclins compared to other organisms. They were divided in different subcategories based on functional similarity with their mammalian counterparts. *Arabidopsis* contains 10-A type, 11 B-type, 10 D-type, and 1 H-type (Vandepoele et al., 2002). Furthermore, other cyclin-related genes were identified classified as P-type, L-type, T-type and C-type cyclins (Torres Acosta et al., 2004; Wang et al., 2004; Menges et al., 2005), however for many of these their function in cell cycle still has to be demonstrated.

Based on their expression profiles, it is assumed that D-type cyclins act at cell cycle (re-)entry and at G1/S transition, although some studies propose that D-type cyclins might also act at mitosis (Nakagami et al., 1999; Kono et al., 2003). Further, A-type cyclins are supposed to control S/M phases, and B-type cyclins both the G2/M transition and the intra M-phase (Menges et al., 2005).

Plant growth factors like auxins, cytokinins, brassinosteroids, gibberellins and sucrose modulate expression of D-type cyclins (Inze and De Veylder, 2006) and some of them might be key switches triggering hormone responses, as is demonstrated by the cytokinin-independent growth of cell cultures overexpressing CYCD3;1 (Riou-Khamlichi et al., 1999). Due to genome duplication, there must be some extend of functional redundancy which is reflected by the minor effects observed when eliminating expression of a specific D-type cyclin (Swaminathan et al., 2000). Also knock-outs of A-type cyclins showed only marginal phenotypes, such as an increase in DNA ploidy level (Imai et al., 2006). On the other hand, overexpression of A-type cyclins had more severe effects. For instance, overproduction of CYCA3;2 triggers ectopic cell divisions and delays differentiation (Yu et al., 2003), indicating that differentiation requires down-regulation of CDK activity.

Furthermore, *Arabidopsis* contains 2 CDK subunit (CKS) proteins, related to yeast SUC1. They are proposed to act as scaffold proteins of CDK/cyclin complexes. Although their precise function still remains to be resolved, they could influence the interactions between the

kinase complex and the substrates. Different analysis showed interaction of *Arabidopsis* CKS proteins with CDKA;1 and CDKBs (Boudolf et al., 2001; Van Leene et al., 2007). The expression of CKS1 is associated with dividing cells (De Veylder et al., 1997b) as well as with endoreduplication cells (Jacqmard et al., 1999).

CDK inhibition

Cell cycle progression is negatively regulated by different mechanisms. First of all, docking of small proteins to CDK/cyclin complexes negatively affects its activity. Arabidopsis encodes 7 proteins related to the mammalian Kip/Cip inhibitors, known as Kip-related proteins (KRPs). All KRPs share a C-terminal domain involved in binding of CDKs and cyclins and essential for the inhibitory activity. All 7 KRPs interacted *in vitro* with D-type cyclins (Inze and De Veylder, 2006), which was further supported by the fact that the inhibitory effect from overexpression of KRPs was complemented by co-overexpression of D-type cyclins (Zhou et al., 2003). Protein stability of KRPs is determined by the presence of destruction motifs in their N-terminal region, leading to SCF-mediated proteolysis, as discussed below. In response to antimitogenic signals, KRPs can induce cell cycle arrest or delay cell cycle progression. Furthermore, they are important both during regular cell cycle, and in plant development since they trigger the switch from the mitotic to the endoreduplication cell cycle in a dose-dependent manner (Verkest et al., 2005a).

In plants, a second family of cell cycle inhibitor proteins exist, comprising SIAMESE (SIM) and SIAMESE-Related (SMR) proteins. SIM shares a motif with KRPs and a putative cyclin-binding motif, and is expressed throughout the plant, including in meristems, leaf primordia and trichomes. It is a nuclear protein promoting endoreduplication in trichomes by suppression of mitosis. The fluorescence resonance energy transfer (FRET) method revealed interaction of SIM with CDKA;1 and D-type cyclins, and the authors propose that it inhibits mitosis through regulation of CycB1;1 expression by inhibition of these CDKA;1/CycD complexes (Churchman et al., 2006) triggering endoreduplication. Furthermore, SIM, SMR1-5 and EL2, the rice (*Oryza sativa*) homologue of SIM, are upregulated by abiotic and biotic stress, linking cell cycle progression with stress responses (Peres et al., 2007).

Arabidopsis contains also a Wee1 homolog, inhibiting CDKA;1 through phosphorylation on a Tyr residue upon cytokinin deprivation, osmotic stress, or DNA damage (Inze and De Veylder, 2006). However, functional orthologs of Cdc25, the phosphatase responsible for removal of this inhibitory phosphorylation have not been identified yet in *Arabidopsis*.

Proteolysis

Plant cyclins and other cell cycle regulators, such as the inhibitory KIP-related proteins (KRPs), are subject to extensive regulation by proteolysis. This targeted destruction ensures that cell cycle moves in one direction, driving the cycle forward. This proteolysis runs via the ubiquitin-proteasome pathway, during which ubiquitin chains are consecutively added to the target protein, marking it for irreversible destruction by the 26S proteasome. Ubiquitination requires the generation of polyubiquitin chains on target proteins through the combined action of ubiquitin-activating enzymes (E1s), ubiquitin-carrying enzymes (E2s) and ubiquitin-protein ligases (E3s) that bring targets and E2s together. Two related E3 complexes are most intimately dedicated to basic cell cycle control, namely the anaphase-promoting complex (APC) (Capron et al., 2003a) and the Skp1/Cullin/F-box (SCF)-related complex. Target proteins are recognized by specific motifs in their protein sequence such as the D-box, present in A- and B-type cyclins, the KEN-box, GXEN-box, A-box, C-box, O-box and the TEK-box. Non-degradable versions of the cyclins have proven the functional significance of cyclin destruction, causing severe growth retardation and abnormal development (Weingartner et al., 2004).

The APC is a highly conserved E3 formed by at least 11 subunits, although its minimal ligase module comprises only APC2 and APC11. Sequence homology revealed the existence of all APC core vertebrate counterparts in the *Arabidopsis* genome (Capron et al., 2003a). All APC core genes, with the exception of APC3/Cdc27, are unique and elimination can impair female gametophyte development, as was proven for APC6/Cdc16 (Kwee and Sundaresan, 2003) and APC2 (Capron et al., 2003b). In contrast to the APC core subunits, multiple APC activators exist in *Arabidopsis* (6 Cdc20 and 3 CCS52 (Cdh1-related) genes). As these APC activators confer substrate specificity to the APC, the existence of 9 putative activators possibly reflects the need to target the many mitotic cyclins and other substrates present.

The different expression profiles of these APC activators suggest consecutive actions in the plant cell cycle from late G2 till late anaphase by APC^{Cdc20} and from late mitosis till G1/S transition by APC^{CCS52}. Next to mitotic cyclins, the APC targets proteins of the spindle checkpoint complex, at least in other organisms. This complex monitors progress during mitosis, delaying anaphase onset until each chromosome is correctly attached to the spindle. Unattached kinetochores trigger formation of a Mad2/Bub3/BubR1 complex, which in turns inhibits Cdc20 APC activators, thereby preventing degradation of several cell cycle regulators and progression of anaphase (Kimbara et al., 2004). Homologs of these Mad/Bub proteins are identified in *Arabidopsis* showing similar expression profile as the mitotic cyclins, hinting that also in plants, they might be involved in metaphase to anaphase transition.

Stability of D-type cyclins like CYCD3;1 (Planchais et al., 2004) is on the other hand mediated through targeted proteolysis by SCF-related complexes. The PEST-sequences, rich in Proline, Glutamate, Serine, and Threonine, present in many D-type cyclins mark these proteins for targeted proteolysis. Other cell cycle regulators such as KRPs (Verkest et al., 2005b; Kim et al., 2008; Liu et al., 2008) are also degraded through this SCF-pathway. In this SCF complex, substrate recognition is provided by the F-box protein, however it still is not clear which of the 694 F-box proteins of *Arabidopsis* are involved in destruction of cell cycle proteins. Often, these target proteins require phosphorylation by e.g. CDKs prior to destruction.

The RBR/E2F/DP pathway

Despite its long evolutionary distance, both plants and animals use the same highly conserved pathway to control the G1/S transition. This pathway involves the E2F transcription factor and its dimerization partner DP, which are repressed by interaction with Rb. This complex regulates transcription of genes mainly involved in entry into and execution of S phase and cell cycle progression, identified through the presence of the canonical E2F motif (TTTCCCGC) in their promoter sequence. *Arabidopsis* and other dicots contain one Rb-related protein (RBR) interacting with a conserved LxCxE motif of D-type cyclins and with CDKA;1 (Boniotti and Gutierrez, 2001). Knock-outs of RBR showed that this protein controls the arrest of the mature unfertilized megagametophyte (Ebel et al., 2004). Furthermore it regulates differentiation of root stem cells (Wildwater et al., 2005). *Arabidopsis* is

characterized by the presence of 3 typical E2Fs (E2Fa, E2Fb and E2Fc), 3 atypical E2Fs (DEL1-3), and two DPs (DPa and DPb). While E2Fa and E2Fb promote cell division, E2Fc acts as a negative regulator of the E2F-responsive genes, because it lacks a strong activation domain. Experimentally validated E2F targets in *Arabidopsis* comprise MCM3, CDC6, CDT1a, PCNA, RBR, and RNR, while *in silico* analysis identified target genes involved in DNA replication, cell cycle regulation, and chromatin dynamics (Ramirez-Parra et al., 2003; Vandepoele et al., 2005). Binding to the E2F motif requires dimerization of E2Fs and DPs as they contain only one DNA-binding domain. By contrast, the 3 atypical DEL proteins posses 2 DNA-binding domains that allow them to bind the E2F motif in a DP-independent monomeric way. Moreover, these proteins lack a transactivation domain, and as such, they might be involved in negative regulation of E2F-activated promoters. They also do not contain a Rb-binding motif. As DEL1 peaks both at G1/S and G2/M transition and not in S-phase, it might control the temporal expression of E2F target genes. Furthermore, it plays an important role in DNA endoreduplication, through activation of CCS52A2 expression (Lammens et al., 2008). CCS52 on his turn, probably destroys mitotic cyclins, and as so, the mitotic cell cycle switches to the endoreduplication cycle. During endoreduplication cells undergo iterative DNA replications without any subsequent cytokinesis, leading to increased ploidy levels. DEL3 on the other hand, regulates cell wall biogenesis genes (Ramirez-Parra et al., 2004).

Chapter 4: Identification of the cell cycle interactome of *Arabidopsis thaliana* through a targeted proteomics approach

Jelle Van Leene, Jens Hollunder, Dominique Eeckhout, Geert Persiau, Eveline Van De Slijke, Hilde Stals, Gert Van Isterdael, Sandy Neirynck, Yelle Buffel, Stefanie De Bodt, Steven Maere, Kris Laukens, Anne Pharazyn, Nubia Eloy, Paulo C.G. Ferreira, Lieven De Veylder, Pierre Hilson, Martin Kuiper, Harry Van Onckelen, Dirk Inzé, Erwin Witters, and Geert De Jaeger

Abstract

Genome sequence analysis has predicted vast numbers of cell cycle regulatory genes in the model plant *Arabidopsis thaliana*. For many of these, microarray experiments or mutant analysis have shown their transcriptional regulation or regulatory role during the cell cycle. To further extent their functional annotation, we isolated protein complexes by tandem affinity purification for 102 cell cycle related proteins. The resulting protein interaction network among 394 proteins contains 866 interactions, from which 82% has never been described before. Based on a computational analysis, gene networks involved in cell cycle phase transition, DNA replication, and mitosis are presented, together with a list of new candidate cell cycle proteins. This is the first cell cycle interactome for higher eukaryotes mapped by tandem affinity purification of protein complexes and mass spectrometry analysis. It provides insight in the basic cell cycle machinery and serves as a guide for the investigation of other protein networks by complex purification in plants.

Manuscript in preparation

4.1 Introduction

Regulation of cell division is of pivotal importance for plant growth and development. The basic underlying mechanisms of cell division are conserved among all eukaryotes. However, mainly due to their sessile lifestyle, plants have evolved unique features, including an indeterminate mode of post-embryonic development allowing rapid integration of environmental and internal signals with cell cycle regulation. This may explain why plants have evolved a high number of cell cycle regulators. The core key players regulating cell cycle progression are Ser/Thr kinases, known as cyclin-dependent kinase (CDK). Substrate specificity of these kinases is provided by the cyclin partner (Nurse, 1994). Other cell cycle proteins regulate directly the activity of these CDK/cyclin complexes through binding or modification (Inze and Veylder, 2006).

When the genomic sequence of the model plant *Arabidopsis thaliana* became available, sequence homology analysis revealed the existence of 14 CDKs and 15 CDK-like genes, 10 CDK-interacting or modifying proteins, an unexpected high number of 38 cyclins, and 9 proteins showing homology to the E2F/DP/RB family, that control the start of DNA replication (Vandepoele et al., 2002; Menges et al., 2005). They were defined and predicted as the 'core' cell cycle machinery of *Arabidopsis thaliana*, regulating the progression through the cell cycle. Microarray analysis confirmed that the majority of these showed a cell cycle phase-dependent expression profile (Menges et al., 2005). The same analysis further discovered new candidate cell cycle genes, peaking either at S-phase or at the G2/M boundary, including homologues of the mitotic checkpoint control proteins MAD2 and BUB1, and homologues of the anaphase promoting complex (APC) activators. The *Arabidopsis* homologues of the APC were also identified by sequence homology searches (Capron et al., 2003a). The APC is an E3 ubiquitin ligase, targeting mitotic proteins for irreversible destruction by the 26S proteasome, steering progression through the cell cycle, as shown for yeast and animal cells (Peters, 2006). Finally the *Arabidopsis* DNA replication machinery was also predicted by a similar sequence analysis approach (Shultz et al., 2007). In parallel with the genome wide discovery of cell cycle genes, mutant analysis and other studies confirmed the role of many of these genes in cell division. Based on this knowledge, the first cell cycle

models are being built for plants (Dewitte and Murray, 2003; Inze, 2005; Inze and Veylder, 2006; De Veylder et al., 2007; Francis, 2007).

Now that a substantial amount of cell cycle genes have been cloned, a comprehensive functional study analyzing protein-protein interactions involving these cell cycle genes would offer new and valuable system-wide information. The objective of this work is to build a cell cycle interactome for plants to unravel the interactions and complexes involved in cell division, and to further sustain the involvement of putative cell cycle regulators in cell division. This interactome should allow us to gather additional functional knowledge to further shed light on the regulation of cell division. In addition, it should help us find new cell cycle regulators and to find pathways connecting the cell cycle machinery with other processes. To elucidate the cell cycle interactome we used the tandem affinity purification (TAP) strategy (Rigaut et al., 1999) to isolate cellular complexes around 108 cell cycle proteins. In our previous study, we showed that TAP on plant cells from *Arabidopsis* cell suspension cultures allows the efficient isolation of cell cycle multiprotein complexes (Van Leene et al., 2007). These cell suspension cultures are extremely well suited to study the basic cell cycle machinery in higher eukaryotes: first of all, they rapidly provide an unlimited and cheap supply of proliferating cells, a prerequisite considering that cell cycle proteins often accumulate at low levels. Second, more than 85 % of the predicted cell cycle proteins are expressed in this system (Menges et al., 2005). Third, protein-interactions can be studied in the absence of development, pinpointing at the real basic cell cycle machinery. And finally, in cell culture most cells are actually dividing, in contrast to the *in planta* situation, where most cells are differentiated and the mitotic cell cycle is turned off. Dividing cells are only found in small parts of the plant, mainly in the shoot and root apical meristems, in the leaf, flower and lateral root primordia and in the cambium. Although yeast cultures are also an ideal system to study the basic cell cycle machinery - as they are very cheap and easy to grow and as they allow easy synchronization - it is a lower eukaryote and there will be huge differences in certain aspects of cell cycle regulation compared to higher eukaryotes. Cell cultures derived from mammalian organisms on the other hand could offer very valuable information as a wide variety of human diseases are the consequence of a perturbed cell cycle. However, these mammalian cells often grow in monolayers requiring rich and

expensive growth sera generating limitations in biomass needed to perform TAP. So, our system is superior to the previously mentioned, as it is derived from a higher eukaryote, as biomass is not a limiting factor and as it is synchronizable."

Here, we provide a protein interaction map centered towards cell division for *Arabidopsis thaliana*, the model organism for higher plants. This cell cycle interactome contains 866 interactions among 394 proteins, of which 82 % were not reported before.

4.2 Results

4.2.1 A high-confidence protein interaction map centered towards cell cycle

To shed light on the cell cycle interactome, 108 proteins were tagged: 73 'core' cell cycle regulatory proteins (Vandepoele et al., 2002; Menges et al., 2005; Peres et al., 2007), 4 mitotic checkpoint proteins (Menges et al., 2005), 8 anaphase promoting complex (APC) subunits and 6 APC activators (Capron et al., 2003a), one 26S proteasome subunit (Brukhin et al., 2005), and 10 proteins involved in DNA replication or repair. As proof of concept, 6 proteins were chosen for reverse TAP experiments (Table 1).

Table 1: Overview of bait proteins used to elucidate the cell cycle interactome of *Arabidopsis thaliana*. Topology = C- or N-terminal TAP fusions. Tag refers to the applied tag being either the traditional TAP tag developed for *Saccharomyces cerevisiae*, or an optimized dual affinity tag (GS) (Van Leene et al., 2008). Expression was indicated as (+) if the TAP fusion protein could be detected by western blot analysis. The total number of purifications performed per bait is shown, with a minimum of 2 experiments.

Bait	Locus	Category	Topology	Tag	Expressed	Purifications
CDKA;1	AT3G48750	core cell cycle	C + N	TAP	+	11
CDKB1;1	AT3G54180	core cell cycle	C + N	TAP	+	4
CDKB1;2	AT2G38620	core cell cycle	C	TAP	+	2
CDKB2;1	AT1G76540	core cell cycle	N	TAP	+	2
CDKB2;2	AT1G20930	core cell cycle	C	TAP	+	4
CDKC;1	AT5G10270	core cell cycle	C	TAP	+	2
CDKC;2	AT5G64960	core cell cycle	C	TAP	+	4
CDKD;1	AT1G73690	core cell cycle	C	TAP	+	2
CDKD;2	AT1G66750	core cell cycle	C + N	TAP + GS	+	10
CDKD;3	AT1G18040	core cell cycle	C	TAP	+	2

CDKE;1	AT5G63610	core cell cycle	N	TAP	+	2
CDKF;1	AT4G28980	core cell cycle	C	TAP	+	3
CDKG;1	AT5G63370	core cell cycle	C	GS	+	2
CDKG;2	AT1G67580	core cell cycle	C	TAP	+	2
CKS1	AT2G27960	core cell cycle	C	TAP + GS	+	10
CKS2	AT2G27970	core cell cycle	C	TAP	+	2
CYCA1;1	AT1G44110	core cell cycle	C	TAP	+	2
CYCA1;2	AT1G77390	core cell cycle	C	TAP	-	
CYCA2;1	AT5G25380	core cell cycle	N	TAP	+	2
CYCA2;2	AT5G11300	core cell cycle	C	TAP	+	4
CYCA2;3	AT1G15570	core cell cycle	N	TAP	+	4
CYCA2;4	AT1G80370	core cell cycle	C	TAP	-	
CYCA3;1	AT5G43080	core cell cycle	C	TAP	+	4
CYCA3;2	AT1G47210	core cell cycle	N	TAP	-	
CYCA3;3	AT1G47220	core cell cycle	N	GS	+	2
CYCA3;4	AT1G47230	core cell cycle	C	TAP	+	4
CYCB1;1	AT4G37490	core cell cycle	C	TAP + GS	+	4
CYCB1;2	AT5G06150	core cell cycle	C	TAP	+	4
CYCB1;3	AT3G11520	core cell cycle	C	TAP	+	2
CYCB1;4	AT2G26760	core cell cycle	C	TAP	+	2
CYCB2;1	AT2G17620	core cell cycle	N	TAP	+	2
CYCB2;2	AT4G35620	core cell cycle	C	TAP	+	5
CYCB2;3	AT1G20610	core cell cycle	C	TAP	+	4
CYCB2;4	AT1G76310	core cell cycle	C	GS	+	2
CYCB2;5	AT1G20590	core cell cycle	C	TAP	+	2
CYCB3;1	AT1G16330	core cell cycle	N	TAP	+	4
CYCD1;1	AT1G70210	core cell cycle	C	TAP	-	
CYCD2;1	AT2G22490	core cell cycle	C	TAP	+	6
CYCD3;1	AT4G34160	core cell cycle	N	TAP	+	2
CYCD3;2	AT5G67260	core cell cycle	C	TAP	+	3
CYCD3;3	AT3G50070	core cell cycle	C	TAP	+	4
CYCD4;1	AT5G65420	core cell cycle	C	TAP	+	2
CYCD4;2	AT5G10440	core cell cycle	C	TAP	+	2

CYCD5;1	AT4G37630	core cell cycle	C	TAP	+	2
CYCD6;1	AT4G03270	core cell cycle	C	TAP	+	2
CYCD7;1	AT5G02110	core cell cycle	C	GS	+	2
CYCH;1	AT5G27620	core cell cycle	N	TAP	+	4
CYCT1;3	AT1G27630	core cell cycle	C	TAP	+	2
DEL1	AT3G48160	core cell cycle	N	TAP	+	2
DEL2	AT5G14960	core cell cycle	N	TAP	+	2
DEL3	AT3G01330	core cell cycle	N	TAP	+	2
DPa	AT5G02470	core cell cycle	C	TAP	+	2
DPb	AT5G03415	core cell cycle	N	TAP	+	2
E2Fa	AT2G36010	core cell cycle	C	TAP + GS	+	5
E2Fb	AT5G22220	core cell cycle	N	TAP	+	2
E2Fc	AT1G47870	core cell cycle	C	TAP	+	5
KRP1	AT2G23430	core cell cycle	N	TAP	+	2
KRP2	AT3G50630	core cell cycle	N	TAP	+	6
KRP3	AT5G48820	core cell cycle	N	TAP	+	4
KRP4	AT2G32710	core cell cycle	N	TAP	+	6
KRP5	AT3G24810	core cell cycle	N	TAP	+	2
KRP6	AT3G19150	core cell cycle	N	TAP	+	4
KRP7	AT1G49620	core cell cycle	N	TAP	+	2
RBR	AT3G12280	core cell cycle	N	TAP	+	2
WEE1	AT1G02970	core cell cycle	N	TAP	+	4
Cdc25-like	AT5G03455	core cell cycle	C	TAP	+	2
MAT1	AT4G30820	core cell cycle	N	TAP	+	2
SIM	AT5G04470	core cell cycle	C	TAP	+	4
SMR1	AT3G10525	core cell cycle	C	TAP	+	4
SMR2	AT1G08180	core cell cycle	C	TAP	+	2
SMR3	AT5G02420	core cell cycle	C	TAP	+	2
SMR4	AT5G02220	core cell cycle	C + N	TAP	+	4
SMR5	AT1G07500	core cell cycle	N	GS	+	2
APC2	AT2G04660	APC core	N	TAP	+	2
APC4	AT4G21530	APC core	C	TAP	-	
APC7	AT2G39090	APC core	N	TAP	+	2

APC8	AT3G48150	APC core	N	TAP	+	2
APC10	AT2G18290	APC core	N	TAP	+	4
APC11	AT3G05870	APC core	C	TAP	+	2
CDC16	AT1G78770	APC core	C	TAP	+	2
CDC27B	AT2G20000	APC core	C	TAP	+	2
CCS52A1	AT4G22910	APC activator	N	TAP	+	4
CCS52A2	AT4G11920	APC activator	N	TAP	+	6
CCS52B	AT5G13840	APC activator	N	TAP	+	4
CDC20.1	AT4G33270	APC activator	N	TAP	+	2
CDC20.3	AT5G27080	APC activator	N	TAP	+	2
CDC20.6	AT5G27945	APC activator	C	TAP	-	
RPN1a	AT2G20580	26S proteasome	N	TAP	+	2
CDC6	AT2G29680	DNA replication	N	TAP	+	2
CDC6b	AT1G07270	DNA replication	C	TAP	+	2
CTF8	AT5G52220	DNA replication	N	GS	+	2
ETG1	AT2G40550	DNA replication	C + N	TAP	+	4
MCM6	AT5G44635	DNA replication	C	TAP	+	2
MCM7	AT4G02060	DNA replication	C	TAP	+	2
ORC1a	AT4G14700	DNA replication	N	TAP	+	2
PCNA1	AT1G07370	DNA replication	N	TAP	+	2
RPA2	AT2G24490	DNA replication	N	TAP	+	2
UVH6	AT1G03190	DNA repair	C	TAP	+	2
BUB3	AT1G69400	mitotic checkpoint	N	TAP	+	2
BUB3-like	AT3G19590	mitotic checkpoint	N	TAP	+	2
BUBR1-like	AT2G33560	mitotic checkpoint	N	TAP	+	2
MAD2-like	AT3G25980	mitotic checkpoint	N	TAP	+	3
DL3195C	AT4G14310	reverse	N	TAP	+	2
F27G20.14	AT1G32310	reverse	C	GS	+	2
expressed	AT5G40460	reverse	C	TAP	+	2
expressed	AT1G10690	reverse	N	TAP	+	2
UVI4	AT2G42260	reverse	N	TAP	+	2
UVI4-like	AT3G57860	reverse	N	TAP	+	2

TAP expression cassettes, either N- or C-terminal tagged (or both), were cloned with the Gateway technology and were stably integrated in the genome of the *Arabidopsis* cell suspension culture PSB-D by Agrobacterium-mediated transformation (Van Leene et al., 2007). Expression was driven by the constitutive Tobacco Mosaic Virus 35S promoter, as it was previously shown that this promoter leads to higher complex recovery as compared to endogenous promoters (Van Leene et al., 2007). Fusion proteins compete with the endogenous untagged counterparts for assembly into multiprotein complexes. As our cell culture has a ploidy level of 8C, constitutive expression under control of 35S turned out to be an advantage.

Of the 108 bait proteins, 102 proteins were detected by western blot analysis as a fusion protein to the TAP tag (Fig. 1). At least two independent purifications were performed for each of the 102 expressed proteins, and in total, 303 purifications were done. Purified proteins were separated by one-dimensional gel electrophoresis, in-gel trypsinized and identified via MALDI-TOFTOF by peptide mass fingerprinting or MSMS. Intrinsic to this one-dimensional separation approach, we are often dealing with peptide mixtures from different proteins, so to increase coverage, a variety of search parameters was used. Despite the two successive purifications steps, background levels caused by non-specific binding remained high and variable with the traditional TAP tag when applying plant cell protein extracts. Most of these non-specific co-purifying proteins were determined by mock purifications, and by purifications with extracts from cultures expressing TAP or GS fusions of heterologous GFP or β-glucuronidase. As a first filter, these promiscuous contaminating proteins were systematically subtracted from the hit lists (control identifications; Fig. 1 & Supplementary data Table S1), generating a non-redundant dataset of 1734 interactions among 985 proteins. Next, the dataset was divided in a ‘core’ dataset, containing interactions that were biologically confirmed in at least 2 independent experimental repeats or in the reciprocal purification experiment, and the ‘non-core’ dataset containing the remainder. During the course of the project, identifications were done with different versions of the SNAPS database (Van Leene et al., 2007). So, to obtain more uniform results, all identifications from the core and the non-core dataset were resubmitted to Mascot and identified with the protein sequence repertoire from the latest TAIR database (TAIR8.0).

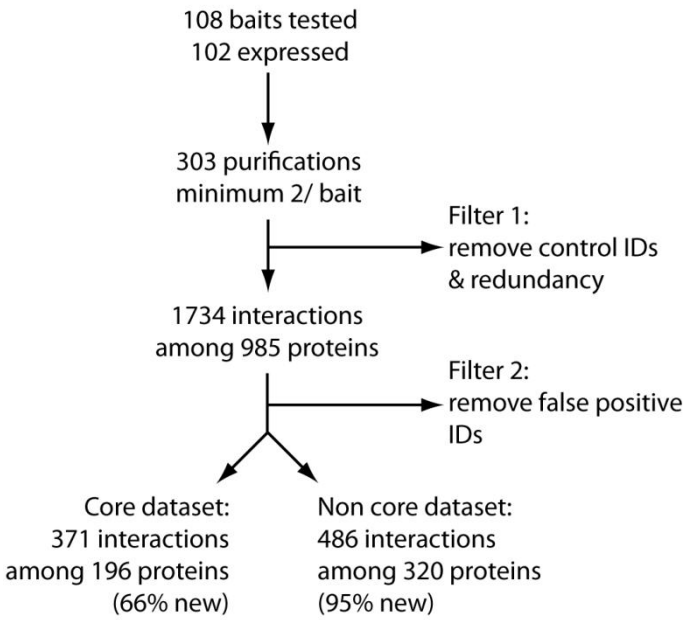


Figure 1. Flow diagram illustrating the process used to generate the core and non-core cell cycle dataset.

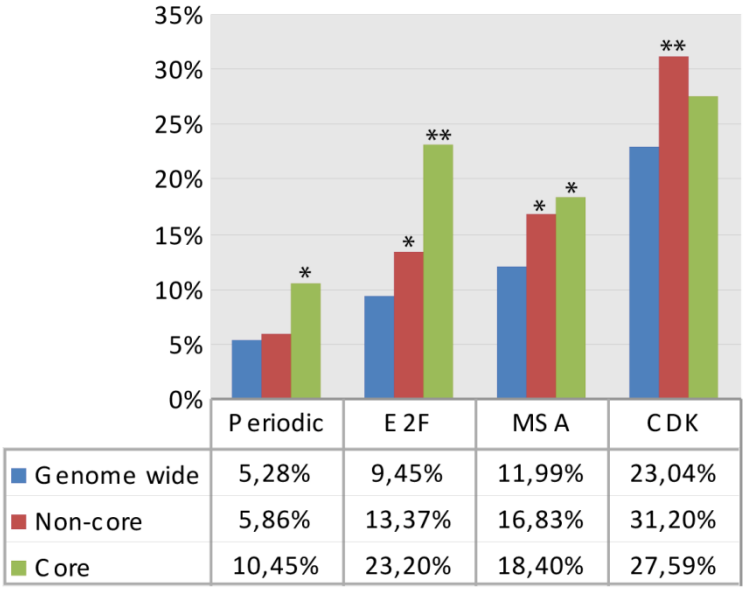
Furthermore, by comparison of mass spectrometry parameters obtained for core and non-core interactions, an additional restriction was implemented to reduce the number of false positive identifications: protein identifications that relied on a peptide mass fingerprint, populated for more than 50% with peptides containing a missed tryptic cleavage, were considered as false positive and discarded. This approach gave rise to a cell cycle interactome of 857 interactions among 393 proteins, divided into a high-confidence core dataset of 371 interactions among 196 proteins, and a non-core dataset of 486 interactions among 320 proteins, meaning that almost half of all identified interactions were reproducible (Fig. 1). Since at least one confirmed interactor could be identified for 87 of the 102 expressed bait proteins, a complex purification success rate of 85 % was reached, which is in line with analogous genome-wide studies of protein complexes in yeast (Gavin et al., 2006).

4.2.2 Quality assessment

To evaluate the quality of the core and non-core dataset, a computational analysis was performed. To assess which GO terms were statistically significant overrepresented among the preys of both datasets, we used the BiNGO tool (Maere et al., 2005). Both datasets were enriched for the GO term 'cell cycle' with respectively a corrected *p*-value of 7.17E-32 and

6.38E-09 (See Supplementary data Table S2), however for the non-core dataset cell cycle was not the top hit, demonstrating that this dataset is more linked to other biological processes. Next, we screened if there was an enrichment for genes with a periodic expression profile during cell cycle. A list of 1258 genes showing cell cycle regulated and cell cycle associated expression was compiled for this analysis (Menges et al., 2003; Jensen et al., 2006). We observed a 2-fold enrichment among the preys of the core dataset (hypergeometric probability distribution *p*-value 0.0113) (Fig. 2).

Subsequently, we looked if there was an enrichment for genes with E2F or M-specific activator (MSA) motifs in their promoter sequence, as it is known that these genes are targets of gene regulatory networks at the G1/S transition or mitosis respectively. These target genes were *in silico* determined by combining transcript expression data and comparative genomics (Vandepoele et al., 2006).



	Periodic	E2F	MSA	CDK
Genome wide	5,28%	9,45%	11,99%	23,04%
Non-core	5,86%	13,37%	16,83%	31,20%
Core	10,45%	23,20%	18,40%	27,59%

Figure 2. Enrichment analysis on the core and non-core datasets. Statistically significant enrichments are indicated with * (*p*-value < 0.05) or with ** (*p*-value < 0.01). An enrichment analysis was performed for genes showing cell cycle-regulated expression (Periodic), for genes containing an E2F or MSA consensus motif in their promoter sequence and for proteins containing a CDK consensus phosphorylation site in their sequence. For the first three analysis, genome wide corresponds to all genes present on the *Arabidopsis* ATH1 genome array of Affymetrix, while CDK phosphorylation sites were determined for all TAIR8.0 entries with the Patmatch tool at TAIR. Proteins that could not be assigned to a specific gene locus were discarded from the analysis.

For both E2F and MSA motifs, significant enrichments were found. The prey-list of the core dataset is 2.5-fold enriched in genes with an E2F motif (*p*-value 4.14E-06) and 1.5-fold in genes with an MSA motif (*p*-value 2.37E-02) (Fig 2). Also the non-core dataset is enriched for genes with an E2F (1.4-fold; *p*-value 4.16E-02) or MSA (1.4-fold; *p*-value 2.56E-02) motif, however less pronounced (Fig 2). The higher enrichments for the core dataset indicate that it is biased towards stable interactions among proteins belonging to the core cell cycle machinery, while the non-core dataset is biased for more transient interactions linking the core cell cycle machinery with other pathways. This is further supported by the fact that the non-core dataset is more enriched (1.35-fold; *p*-value 2.50E-03) for potential CDK substrates, as compared to the core dataset, as assessed by the presence of the CDK consensus phosphorylation site [ST]PX[KR], a known hallmark of CDK substrates (De Veylder et al., 1997a). The fact that these interactions often could not been confirmed is probably due to the transient nature of kinase/substrate-interactions.

Forty-six per cent of interactions in the core dataset and eight per cent in the non-core dataset are between bait proteins, demonstrating the fidelity of both datasets as our bait proteins are supposed to act in a common pathway.

This was further backed up by a GO similarity analysis (Fig. 3). This score takes into account the shortest path of the common ancestor to the root of the GO tree. This means, if two genes have no related molecular function, do not act in a related biological process, or do not locate to a similar subcellular localization, their common GO ancestor will be close to the root, and the GO similarity score will be low. The distribution of the random gene pairs follows the expected exponential decrease, while for the core dataset one could expect a Gaussian curve around a higher score. However, due to the far from complete GO annotation achieved today for Arabidopsis and to the relative small size of our dataset we obtained a somewhat odd projection of the GO similarity distribution. Nevertheless, although we did not observe an exact Gaussian curve, we demonstrate that gene pairs from the core dataset tend to have a higher score compared to the average distribution of gene pairs from 1000 randomized datasets, what can be seen by the higher fraction having a GO similarity score of 7 for biological process, 4 for cellular component and 7 for molecular

function, meaning that gene pairs of the core dataset tend to act in the same biological process, have a similar function or localize together. For the non-core dataset, this pattern is only observed for the cellular component, meaning that proteins of the non-core dataset can indeed meet each other as is expected. However this also strengthens our hypothesis that the non-core dataset is biased for more transient interactions linking the core cell cycle machinery with other pathways.

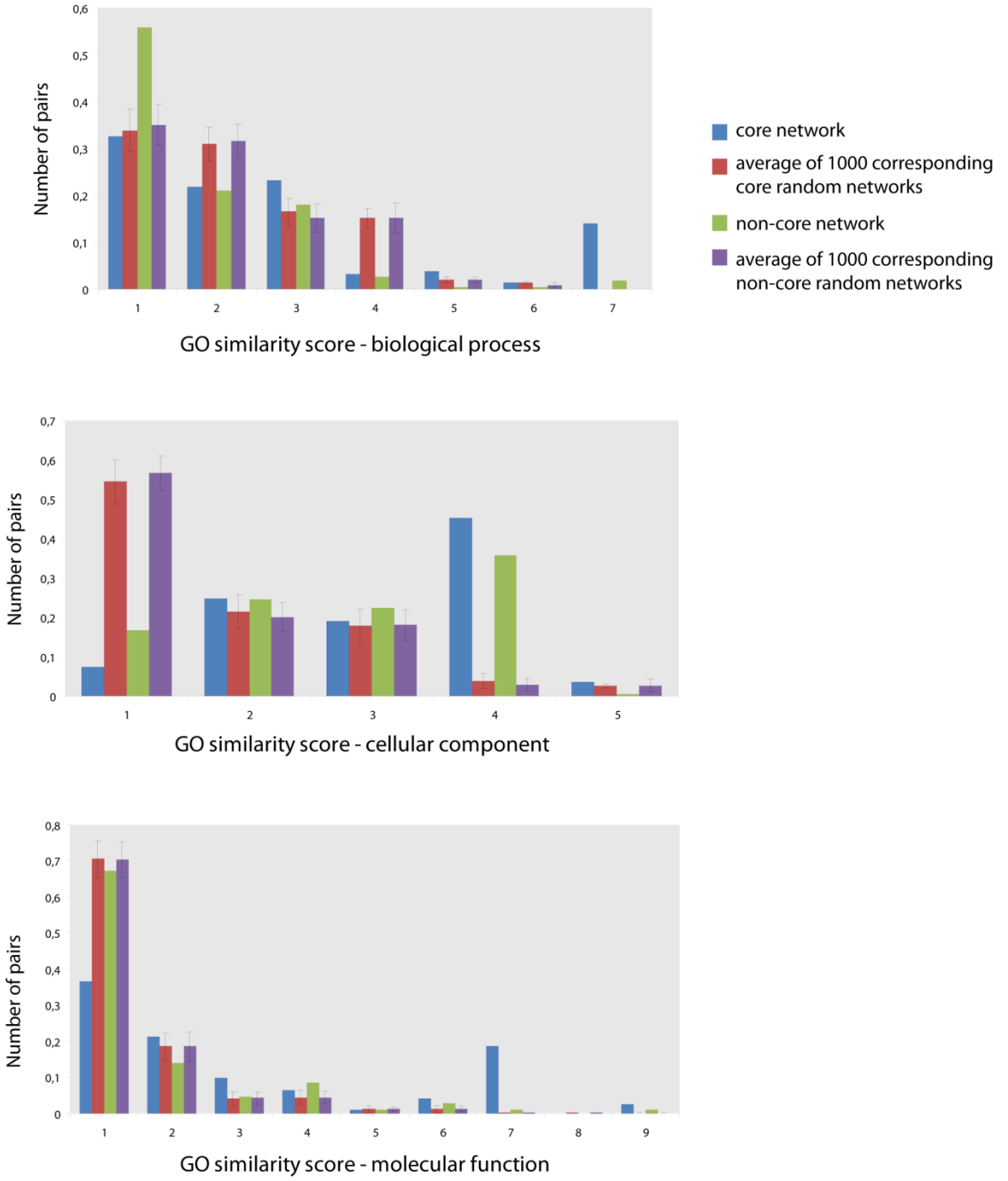


Figure 3. Diagrams showing the distribution of the GO similarity scores of gene pairs of the core network (blue) and the non-core network (green). This is compared to the distribution of the average GO similarity scores of 1000 corresponding networks containing an equal number of randomly chosen gene pairs of the core (red) and non-core (purple) networks. Standard deviations are shown for the average of the random networks. GO terms representing the biological process, cellular component and molecular function were assessed. Proteins that could not be assigned to a specific gene locus were discarded from the analysis.

To assess the novelty of the cell cycle interactome, we screened for the overlap of our datasets with protein-protein interaction databases. Sixty-six per cent of the core dataset interactions are considered new, as they are not documented in TAIR (Huala et al., 2001), InTact (Kerrien et al., 2007), or *Arabidopsis* Reactome (Tsesmetzis et al., 2008), nor predicted at the *Arabidopsis* protein interaction database AtPID (Cui et al., 2008), Reactome (Vastrik et al., 2007) or The Bio-Array Resource (BAR) for *Arabidopsis* Functional Genomics (Geisler-Lee et al., 2007). As these new interactions are not predicted at AtPID, Reactome, or BAR, three databases containing mainly interacting orthologs or 'interologs', most of these new interactions might be plant specific. Furthermore, 84 previously predicted interactions not present in TAIR, InTact or the *Arabidopsis* Reactome, are validated within the core dataset. Finally, of the non-core dataset, 95 % was not reported before, providing together with the core dataset a huge stack of new information.

4.2.3 The quest for new cell cycle proteins

To identify potential new cell cycle-related proteins and regulators in our dataset, we integrated different cell cycle-related features, listed in Table 2. These features involve periodicity during cell division, cell cycle-related promoter motifs, CDK consensus phosphorylation sites, and protein destruction motifs as it is known that targeting cell cycle proteins for destruction is driving progression through the cell cycle.

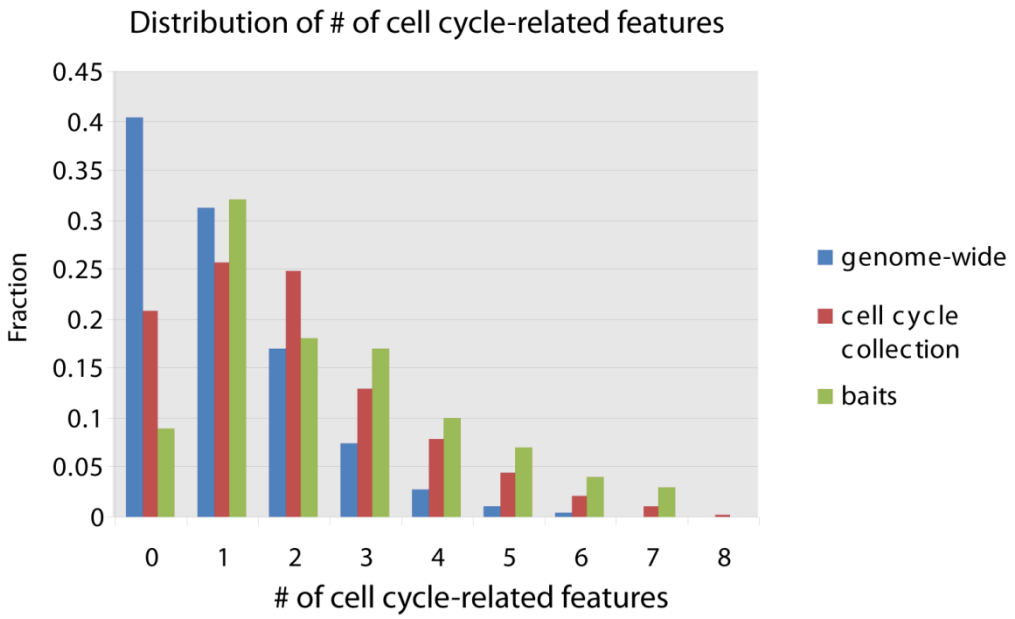
Table 2. Overview of cell cycle-related features used in the computational analysis to search for new cell cycle proteins.

Cell cycle	Category	# of	References
Periodicity	Gene expression	1258	(Menges et al., 2003; Jensen et al., 2006)
MSA-like	Promoter motif	2295	(Vandepoele et al., 2006)
E2Fa-like	Promoter motif	1809	(Vandepoele et al., 2006)
E2F10SPCNA	Promoter motif	2221	(Vandepoele et al., 2006)
OS_motifsIandIIa	Promoter motif	2310	(Vandepoele et al., 2006)
UP1ATMSD	Promoter motif	3738	(Vandepoele et al., 2006)
wrrmGCGn	Promoter motif	2179	(Vandepoele et al., 2006)
CDK consensus site	Phosphorylation motif	6321	(De Veylder et al., 1997a)
[IM]R-tail	Protein sequence motif	116	(Vodermaier et al., 2003; Hayes et al., 2006)

PEST-sequence	Destruction motif	2719	(Rechsteiner and Rogers, 1996)
D-box	Destruction motif	2369	(Capron et al., 2003a)
KEN-box	Destruction motif	410	(Capron et al., 2003a)
GxEN-box	Destruction motif	300	(Castro et al., 2003)
A-box	Destruction motif	1779	(Littlepage and Ruderman, 2002)

We started by testing this hypothesis on known cell cycle genes, which should be enriched for these features. So, first of all, we compiled a collection of 518 cell cycle genes (See Supplementary data Table S3) based on annotation by gene ontology and presence in manuscripts related to cell cycle or DNA replication (Vandepoele et al., 2002; Capron et al., 2003a; Menges et al., 2005; Shultz et al., 2007) as gene ontology is far from complete. The distribution of the number of cell cycle-related features present per gene shows that this collection is enriched for these features compared to the whole gene pool (Fig. 4A). A clear shift between the whole gene pool and the cell cycle compendium was visible at 2 features. This pattern was similar when we considered the original bait list, validating the choice of our baits. In search for new cell cycle proteins, we started from the core and non-core prey list and subtracted the collection of known cell cycle genes. The percentage meeting the criterion of having at least 2 features is shown (Fig. 4B), and all classes are significantly enriched compared to the whole gene pool. By filtering for genes containing at least 2 features, we generated a list of 40 potential new cell cycle genes in the core dataset, including transcription factor-related genes, microtubule-associated proteins, two new APC subunits (see below), proteasome-related proteins, and unknowns (See Supplementary data Table S4). This proofs that also the non-core dataset is a potential source of new cell cycle genes and a list of 83 candidate genes is presented in supplementary data (See Supplementary data Table S5). On the other hand, proteins without or with only one feature should not be disregarded as such, since 46 % of the known cell cycle collection have less than 2 features.

A



B

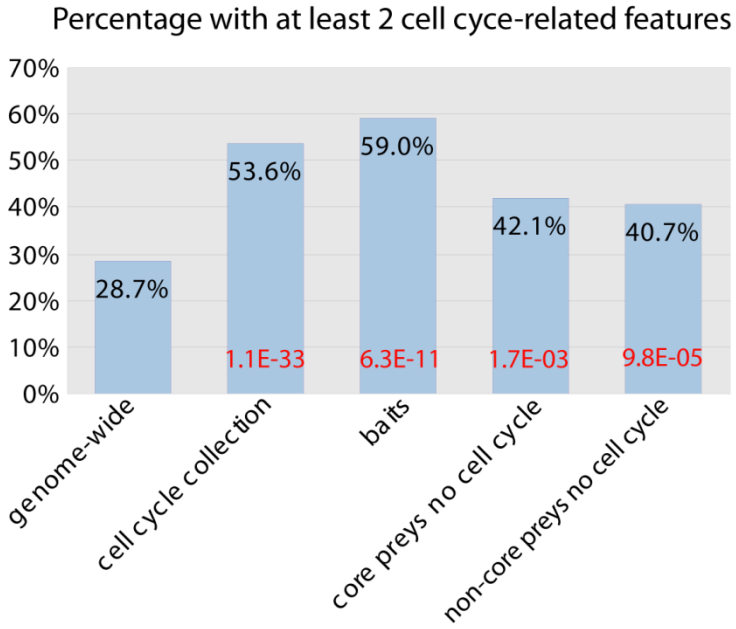


Figure 4. (**A**) Distribution of the number of cell cycle-related features among the whole gene pool (genome-wide), a collection of 518 cell cycle genes, and bait proteins (without reverse baits). (**B**) Percentage of the whole gene pool (genome-wide), the cell cycle collection, bait proteins (without reverse baits), and core and non-core preys (without the cell cycle collection) having at least two features. All classes are significantly enriched for these features compared to the whole gene pool, p-values are shown in red. Proteins that could not be assigned to a specific gene locus were discarded from the analysis.

4.2.4 Data integration, network analysis and network visualization

Previous studies have related protein interactome data with genome-wide transcriptome data (Ge et al., 2003; Li et al., 2004). To investigate to what extend our dataset correlates with transcript expression similarities, we overlapped our interactome data with transcript Pearson Correlation Coefficients (PCC). These PCCs, representing the degree of co-expression of gene pairs, were calculated based on an *Arabidopsis* ATH1 micro-array compendium of 518 experiments focused towards cell cycle or plant growth and development (See Supplementary data Table S6). For the core dataset, we were able to calculate transcript PCCs for 79 % of the interactions. On average, we found a transcript PCC of 0.324, which is dramatically higher compared to the average PCC of 100 randomized datasets (0.016) with an equal number of randomly chosen proteins and interactions (Fig 5). Also for the non-core dataset there is a clear shift to higher PCC values with an average PCC of 0.144 (random 0.016) (Fig. 5).

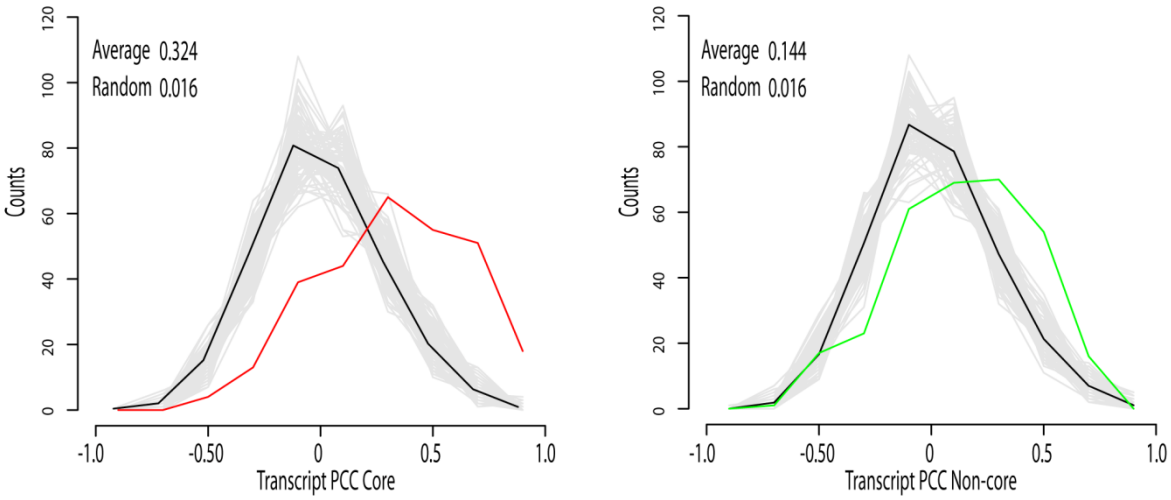


Figure 5. Distribution of the transcript PCC of the core network (red curve) and the non-core network (green curve), compared to the average distribution (black curve) of 100 random networks (grey curves) containing an equal number of interactions and genes (red line). Proteins that could not be assigned to a specific gene locus were discarded from the analysis. Counts = absolute number of gene pairs.

These PCCs can be used to assign confidence to new interactions, as interactors with strong expression correlation are often part of a common molecular assembly (Gunsalus et al., 2005). However, absence of expression correlation does not necessarily suggest a false

positive interaction. Such interactions may exhibit a different relationship, e.g. a regulatory interaction.

The cell cycle interactome data is visualized as an undirected weighted network graphs according to the spoke model (Fig. 6), whereby each bait is assumed to interact with each of its observed prey proteins, and is provided as a Cytoscape file in the supplementary data. Transcript expression profile similarities (PCCs) are integrated into the edge color of the interactome. Furthermore, baits and new candidate cell cycle proteins are highlighted, and novel interactions can be easily discriminated from known interactions, and core from non-core interactions.

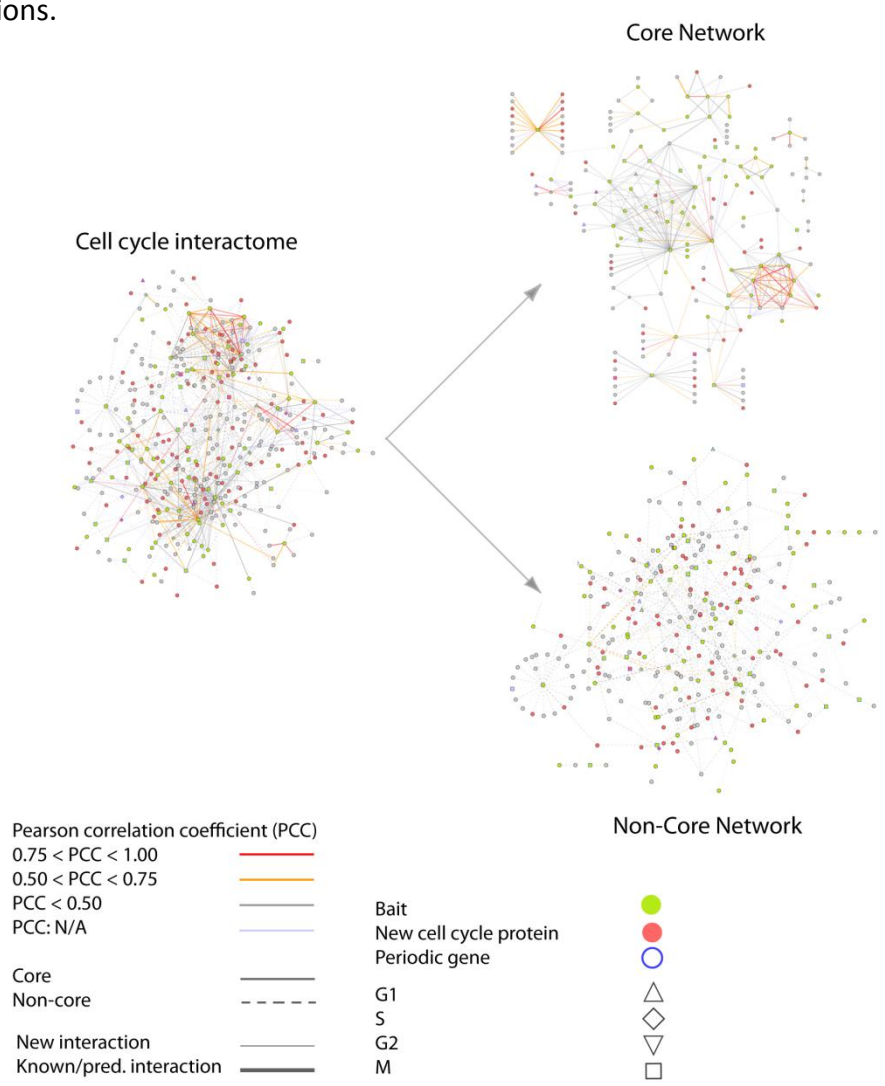


Figure 6. Undirected weighted network graphical views created in Cytoscape representing the whole cell cycle interactome, the core network, and the non-core network. The legend corresponding to the networks is shown.

Similar to other biological networks, the cell cycle interactome network exhibits small-world and scale-free properties, as the degree distribution follows a power-law and as the shortest-path length follows a normal distribution (Fig. 7). The average shortest-path length calculated here is relatively small comparing with the size of the network. This property is usually referred to as a small-world effect.

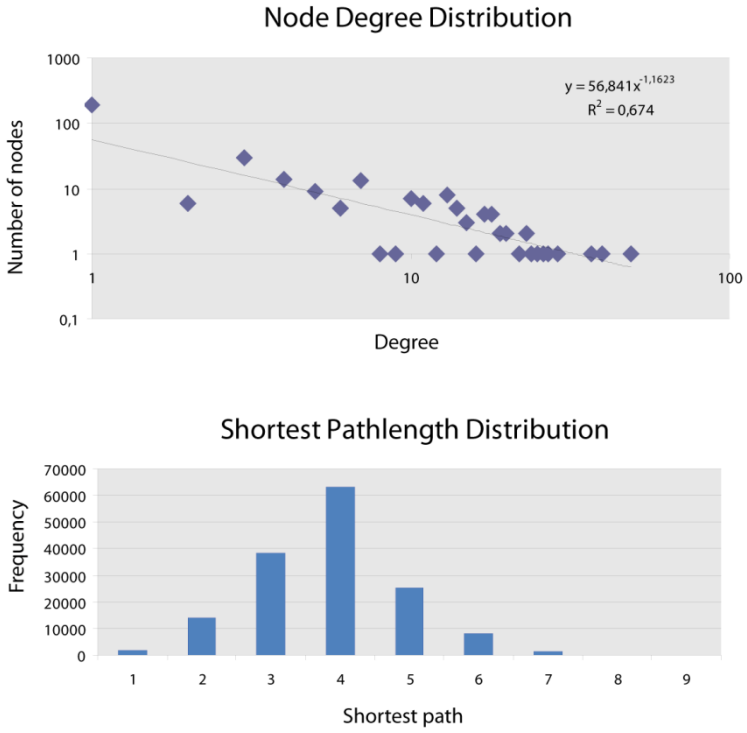


Figure 7. Plot showing the degree distribution and the shortest-path distribution of the whole network (core + non-core). The degree distribution represents the probability distribution of the degrees over the whole network, with the degree being the number of connections it has to other nodes. A path in a graph is a sequence of nodes and edges such that a node belongs to the edges before and after it and no nodes are repeated. The path length is the number of edges in the path.

4.2.5 Birds eye view on the cell cycle interactome

The biologically most important gene networks were extracted from the core and non-core networks, and discussed in the following paragraphs. For the legend with the discussed subnetworks, we refer to Figure 6. As this discussion deals only with a selection of all interactions, the entire interactome can be viewed in the Cytoscape and Excel files (See supplementary data).

The core CDK/cyclin complexes and potential substrates

Progression through the cell cycle is governed by phosphorylation of a plethora of substrates, triggering transition from one cell cycle phase into the next. Key players in these events are cyclin-dependent kinase complexes, consisting out of a catalytic CDK subunit and a regulatory cyclin providing specificity to the CDK complex. Their activity can be further regulated both at the transcriptional level (Menges et al., 2005), and post-translational through phosphorylations, or by binding with inhibitory or scaffold proteins.

The *Arabidopsis* ortholog of yeast cdc28 and mammalian CDK1 and CDK2 is CDKA;1, as it contains the PSTAIRE hallmark in its cyclin-binding domain. Next to CDKA;1, 4 plant-specific B-type CDKs exist in *Arabidopsis*. In contrast to the constant expression levels of CDKA;1, expression of B-type CDKs fluctuates during the cell cycle: while B1-type CDKs are expressed from late S-phase till the end of G2, B2-type CDKs peak mainly at the G2/M boundary (Menges et al., 2005).

CDKA;1 co-purified with all tested D-type cyclins (Fig. 8). These cyclins are upregulated mainly in response to nutrients and other proliferative signals, and have been shown to play a role at the re-entry of the cell cycle, at the G1-S transition (De Veylder et al., 2007) and perhaps also at G2-M transition (Kono et al., 2003; Dewitte et al., 2007). Interaction between a D-type cyclin and a B-type CDK was reported only once before (Kono et al., 2003). We found association of CDKB1;1 with CycD4;1. However, as this interaction was not confirmed, it may reflect phosphorylation of CDKB1;1 by CDKA;1/CycD4;1 complexes regulating its activity. Combining our interactome data with expression data (Menges et al., 2005) we speculate that at cell cycle re-entry and early in G1-phase, CDKA;1 binds CycD3;3 and CycD5;1. Further on in G1-phase and at the G1/S checkpoint CDKA;1 binds a variety of D-

type cyclins, such as CycD4;1, CycD4;2, CycD3;1, CycD6;1 and CycD7;1. During S-phase, CDKA;1 interacts with A3-type cyclins. The other A- and B-type cyclins, of which most possess a similar expression profile at the G2/M-checkpoint, bind with B-type CDKs. B1-type cyclins associate with B2-type CDKs, while the remainder A- and B-type cyclins preferentially bind B1-type CDKs. All CDK/cyclin core complexes co-purify at least one of the 2 scaffold CDK-subunit (CKS) proteins. Despite that mitotic cyclins and CDKBs are highly co-expressed with CKS2, they also interacted with CKS1.

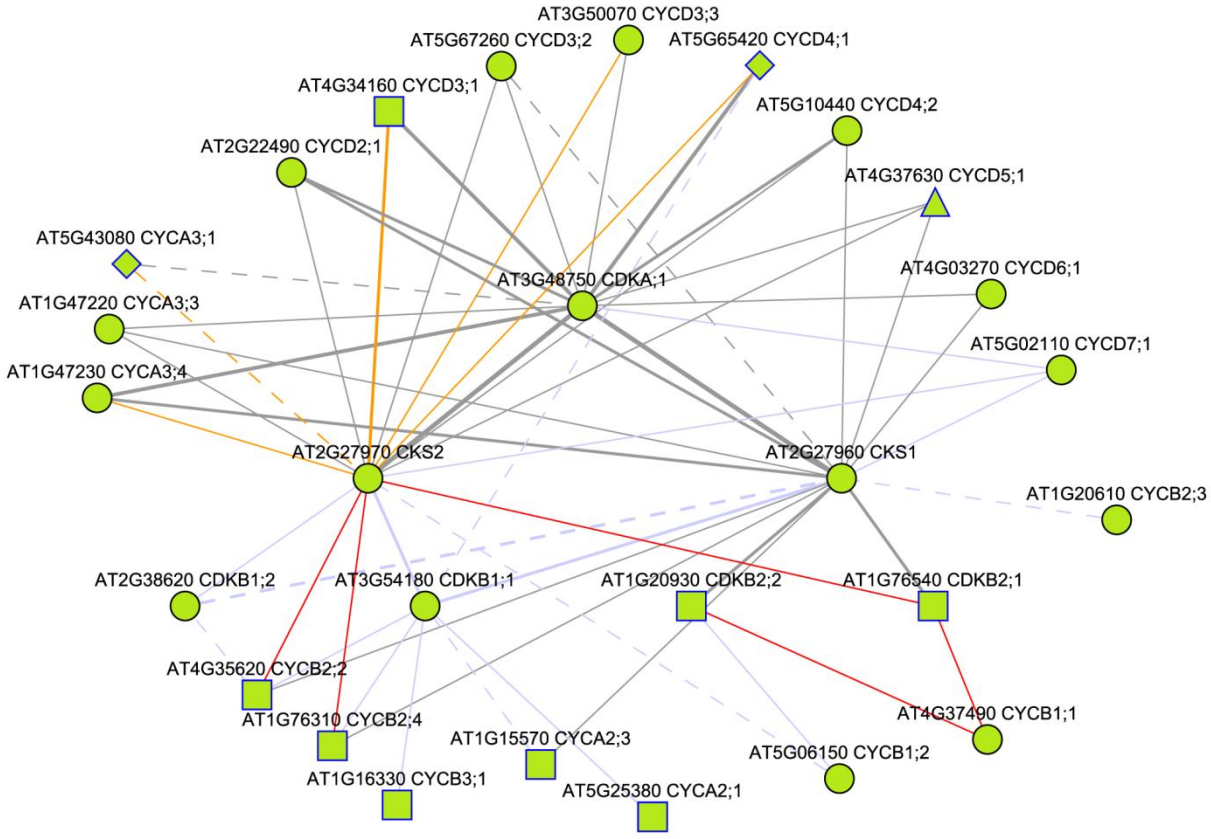


Figure 8 . The core CDK/cyclin subnetwork

Next, we discuss some interesting interactions identified with these core CDKs and cyclins (Fig. 9). As predicted (Geisler-Lee et al., 2007), CDKA;1 is present as a highly connected hub in the core network, interacting with some interesting potential substrates. First of all, CDKA;1 co-purified AT4G14310, a protein of unknown function. This protein was further present in complexes with CKS1, CKS2, CycA3;1, CyCA3;4 and KRP2 and the reverse purification confirmed interaction with CDKA;1 and CKS2 and revealed interaction with the plant-specific kinesin motor protein KCA2, a known substrate of CDKA;1 (Vanstraelen et al.,

2004). As KCA2 is supposed to be involved in division plane determination, AT4G14310 might be involved in the same pathway. Furthermore, CDKA;1 co-purified with RPN1a, a regulatory subunit of the 26S proteasome complex, possibly reflecting cell cycle regulation of the 26S proteasome. With this RPN1a subunit we pulled down 19 proteasome-related proteins. CDKA;1 further interacted with 3 proteins (1 phosphoglucomutase and 2 UDP-glucose 6-dehydrogenases) from the UDP-xylose biosynthesis pathway, coupling cell cycle regulation with cell wall synthesis (Seifert, 2004). With 3 A-type cyclins, we pulled down a DNA repair protein, containing a CDK consensus motif, as a putative substrate. In the non-core dataset, some interesting proteins, possessing a CDK consensus phosphorylation motif and involved in chromatin remodeling were identified with different cyclins: CHR17, an E2F-upregulated ISWI protein (Huanca-Mamani et al., 2005) interacted with CycD3;2 and CycD5;1. CHC1 associated with CycA1;1, CycD7;1, CycB2;3 and CKS2, and BRAHMA, a SWI/SNF chromatin remodeling ATPase involved in the formation and/or maintenance of boundary cells during embryogenesis (Kwon et al., 2006), was identified with CycB1;3 and CycB2;3. B-type cyclins, upregulated during G2 and peaking at G2/M transition, are involved in mitosis. CycB1;3 interacted with the spindle pole body component 98 and γ-tubulin, 2 proteins involved in microtubule (MT) nucleation. These proteins co-localize at nuclear membranes during G2-phase and are involved in assembly of the preprophase band, a plant specific structure required for polarity determination during cell cycle (Erhardt et al., 2002). The interaction with CycB1;3 makes sense, as it is proposed that activation of MT nucleation sites and coordinated regulation of the MT assembly would be controlled by cell cycle and/or developmental signals. CycB2;3 associated with two other interesting proteins, Bonsai 2, involved in repression of cell death and promotion of cell growth (Yang et al., 2006) and NFD5, a protein involved in fusion of polar nuclei during gametophyte development (Portereiko et al., 2006).

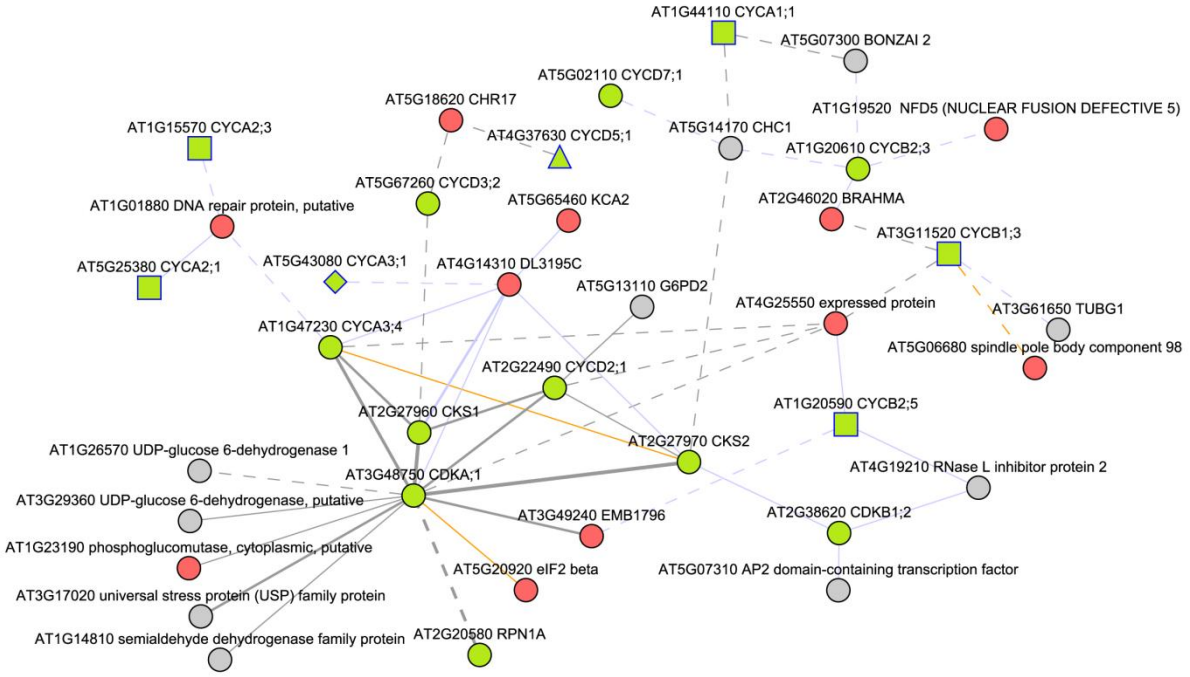


Figure 9. Putative CDK/cyclin substrates

Negative regulation of CDK/cyclin complexes

Cell cycle progression is negatively regulated by docking of small proteins to CDK/cyclin complexes. *Arabidopsis* encodes 7 proteins related to the mammalian Kip/Cip inhibitors, known as Kip-related proteins (KRPs). In response to antimitogenic signals, they can induce cell cycle arrest or delay cell cycle progression. Furthermore, they are important both during regular cell cycle, and in plant development since they trigger the switch from the mitotic to the endoreduplication cell cycle in a dose-dependent manner (Verkest et al., 2005a). Here we confirm that all KRPs, except KRP1, interact with both CDKA;1 and D-type cyclins (Fig. 10). With KRP1, we only found CDKA;1 and no cyclin. We did not observe proposed associations between KRPs and B-type CDKs or A-type cyclins (Verkest et al., 2005a). Importins often co-purified with the nuclear localization signal-containing KRPs, supporting the importance of the regulation of their subcellular localization for their activity. The KRP subnetwork further contains several transcription factors, some proteins of unknown function not only found with KRPs but also with D-type cyclins, the chromatin remodeling protein CHC1, and Decoy, a callose synthesis protein involved in cell plate formation during cytokinesis, that is upregulated in the *HUB1* (Histone Ubiquitination 1 protein) mutant (Fleury et al., 2007).

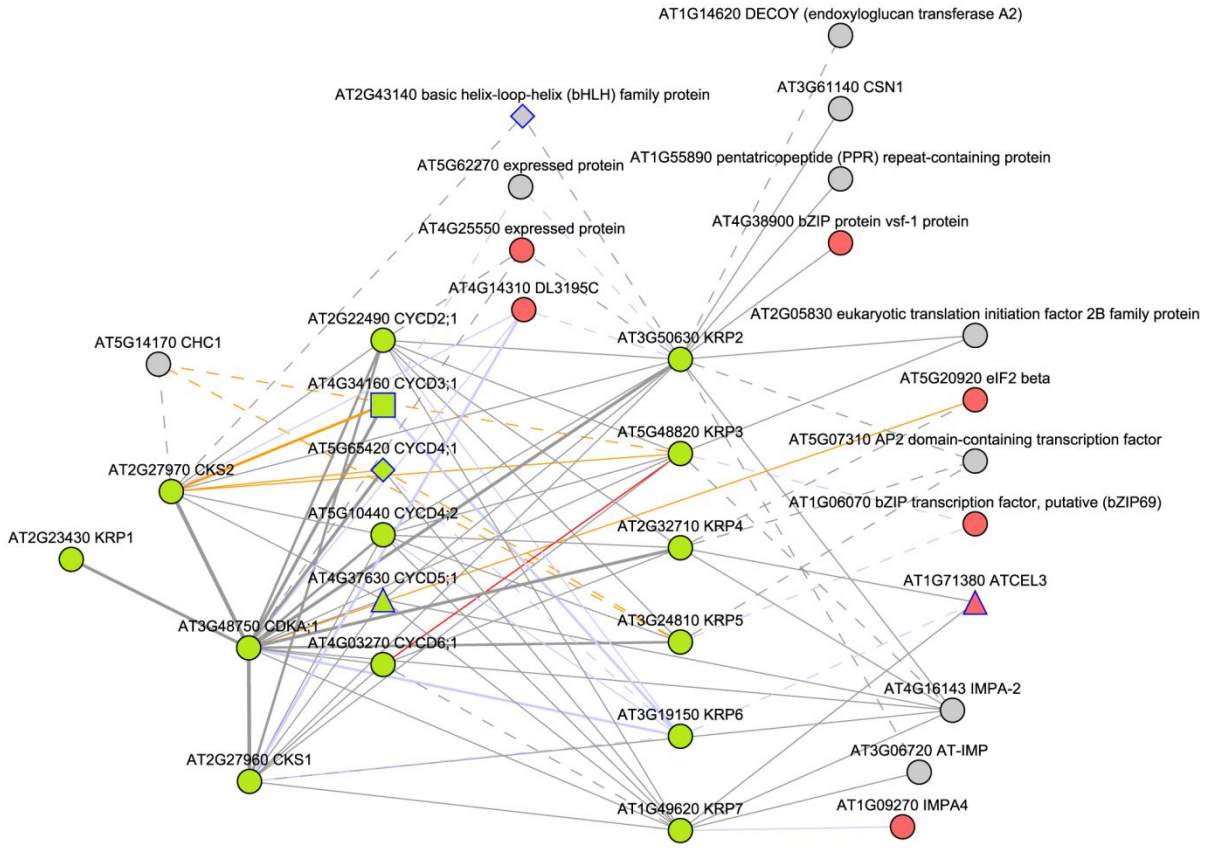


Figure 10. Negative regulation of CDK/cyclin complexes by KRPs

In plants, a second family of cell cycle inhibitor proteins exist, comprising SIAMESE (SIM) and SIAMESE-Related (SMR) proteins. SIM shares a motif with KRPs and a putative cyclin-binding motif, and is expressed throughout the plant, including in meristems, leaf primordia and trichomes. It is a nuclear protein promoting endoreduplication in trichomes by suppression of mitosis. The fluorescence resonance energy transfer (FRET) method revealed interaction of SIM with CDKA;1 and D-type cyclins, and the authors propose that it inhibits mitosis through regulation of CycB1;1 expression by inhibition of these CDKA;1/CycD complexes (Churchman et al., 2006). In our dataset however, SIM co-purifies CDKB1;1 and not CDKA;1, so endoreduplication may be triggered directly by inhibiting mitotic CDKB/cyclin complexes (Fig. 11). Furthermore, SIM, SMR1-5 and EL2, the rice (*Oryza sativa*) homologue of SIM, are upregulated by abiotic and biotic stress, linking cell cycle progression with stress responses. Protein complexes purified via $p13^{Suc1}$ were not inhibited by EL2, postulating that EL2 and CKS may compete for the same binding place (Peres et al., 2007). However, according to our

interactome, in *Arabidopsis*, SIM and all tested SMRs, except SMR2, co-purified at least one CKS protein.

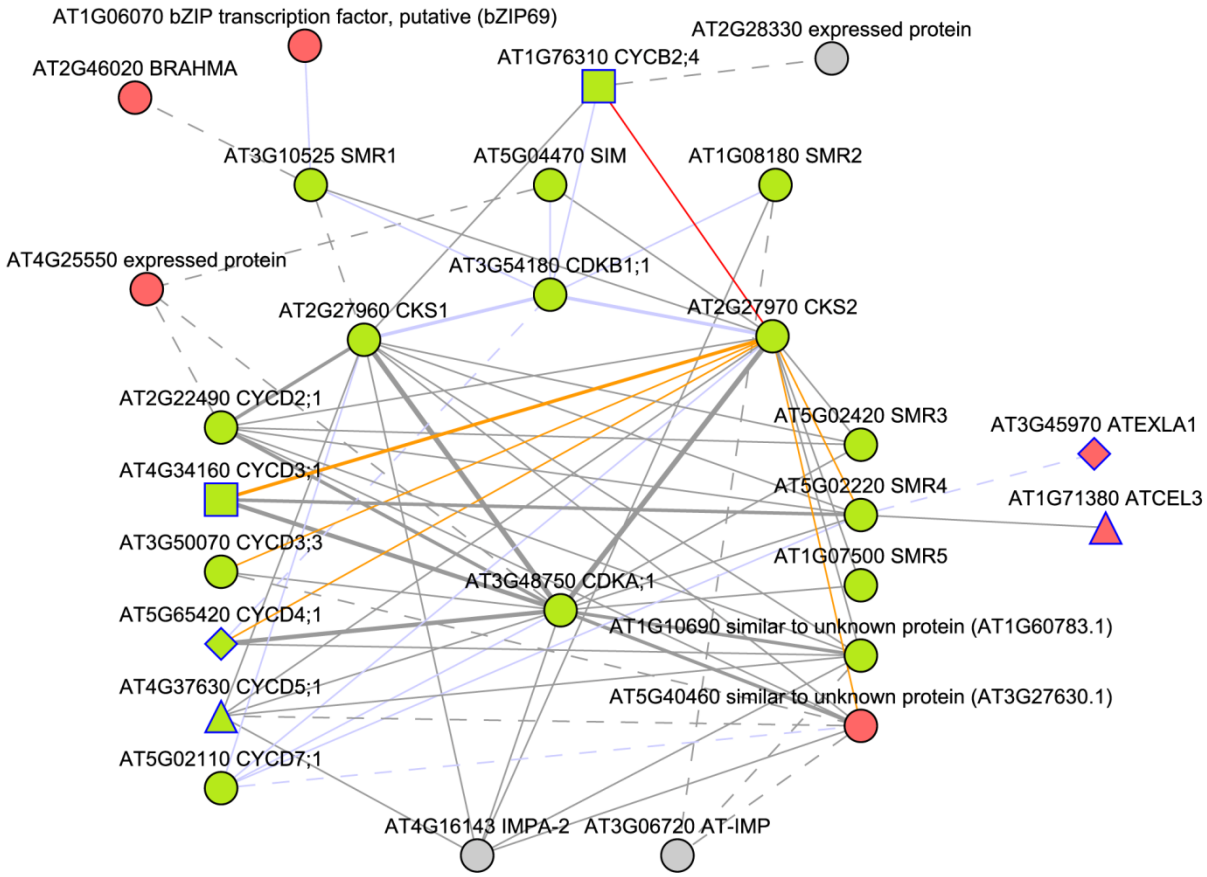


Figure 11. Negative regulation of CDK/cyclin complexes by SIM and SMRs

Next to SIM, also SMR1 and SMR2 associate with CDKB1;1, and the CDKB1;1 interactor CycB2;4 binds AT2G28330, an additional member of the SMR family, discovered on basis of sequence homology (Lieven De Veylder, personal communication). In contrast to SIM, SMR1-2 and AT2G28330, SMR3-5 bind CDKA;1 and D-type cyclins. Besides, with CDKA;1 and some D-type cyclins as bait, we picked up 2 additional members of the SMR clan (Lieven De Veylder, personal communication), AT5G40460 and AT1G10690, and reverse purifications confirmed these interactions. As AT5G40460 was almost 20-fold induced in plants overexpressing E2FxDPa (Vandepoele et al., 2005), it may inhibit CDKA;1/CycD complexes during S-phase preventing re-initiation of DNA replication. Like with the KRPs, we also found nuclear import proteins with the SMRs. SMR1 further co-purified the chromatin remodeling protein BRAHMA and bZIP69, a transcription factor also found with KRP3 and KRP5.

Finally, CDKA;1 can be negatively regulated through inhibitory phosphorylations by Wee1 kinase upon DNA stress (De Schutter et al., 2007), however under the favorable growth conditions of our cell suspension culture, we did not observe this interaction.

Positive regulation of CDK/cyclin complexes

For full activity CDKs require, next to cyclin binding, phosphorylation of a threonine residue within the T-loop by CDK activating kinases (CAK). The *Arabidopsis* genome encodes 4 CAKs, namely 3 D-type CDKs (CDKD;1, CDKD;2 and CDKD;3), homologous to human CDK7, and 1 cyclin-independent CAK-activating kinase (CAKAK) CDKF;1. Here we show that both CDKD;2 and CDKD;3 form a trimeric complex with CycH;1 and the CAK assembly factor MAT1 (Fig. 12). Like in rice (Rohila et al., 2006), CDKD;2 is also part of the basal TFIIH complex involved in transcription and DNA repair, as 3 members co-purified (UVH6/XPD, AT1G55750 and AT4G17020). In this complex, CDKD;2 activates transcription through phosphorylation of the C-Terminal Domain (CTD) of RNA polymerase II. With UVH6 and MAT1 as baits, we confirmed interaction with CDKD;2 and extended the TFIIH complex with two additional proteins (General TFIIH2 and AT1G18340). Interestingly, CDKD;1 is linked to the COP9 signalosome and CycH;1 associated with a chromatin remodeling protein and with nucleoside diphosphate kinase 1, an enzyme involved in nucleotide homeostasis. CDKD;2 co-purified proteins involved in nucleotide biosynthesis, namely 3 ribose-phosphate pyrophosphokinases. More upstream, the monomeric CAKAK CDKF;1 activates CDKD;2 and CDKA;1 in a cyclin-independent manner, however we could not detect the interaction with CDKA;1, nor did we identify the known interaction between CDKA;1 and CDKD;2, most likely because of its transient character. On the other hand, CDKF;1 also binds CDKG;2. The G-type CDK class consist out of 2 members in *Arabidopsis*, and is homologous to the human cytokinesis-associated p58 galactosyltransferase protein. Here we discovered CycL1, a cyclin with a SR-like splicing domain (Forment et al., 2002), as the regulatory cyclin partner of both G-type CDKs, validating the clustering of CycL1 with CDKG;2 in a tissue-specific gene expression analysis (Menges et al., 2005). As CDKG;2 associated with CDKF;1, their activity is probably regulated during cell division, and both core and non-core interacting proteins hint for a function in regulation of transcription and splicing.

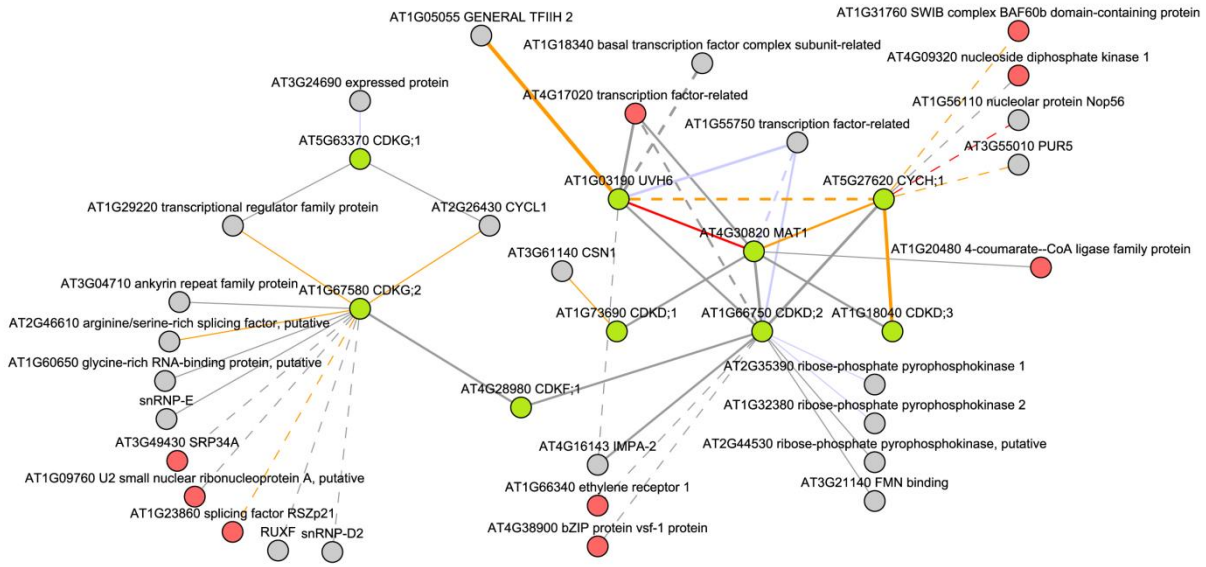


Figure 12. Positive regulation of CDK/cyclin complexes

DNA replication

At the G1/S boundary, proteins of the E2F/DP/RB pathway activate complex gene regulatory networks, inducing transcription of genes mainly involved in nucleotide synthesis, DNA replication and DNA repair. This pathway is extremely well conserved in eukaryotes and comprises 6 E2F factors (E2Fa-c and DEL1-3), 2 dimerization proteins (DPa-b) and one Retinoblastoma-related protein (RBR) in *Arabidopsis*. Activity of E2F/DP complexes is repressed through binding with RBR, regulating normal cell proliferation, endoreduplication and differentiation. Upon hyperphosphorylation of RBR by CDK/cyclin complexes, RBR looses its affinity for E2F/DP, allowing transcription of E2F-target genes. We demonstrate that E2Fa and E2Fb can associate both with DPa and DPb, and that all E2F and DP proteins co-purify RBR (Fig. 13). Similar to E2Fc, the 3 DEL proteins, miss the trans-activation domain and go in competition for E2F promoter binding sites. Atypically, they posses 2 DNA-binding motifs, making binding of E2F-target genes possible in a monomeric DP-independent manner, and they lack a RB-binding motif. This was validated in our interactome as we did not pull down DPs nor RBR with DEL proteins. On the contrary, DEL-purified complexes have some proteins in common with E2Fa or DPb, like the RNAse L inhibitor, an endogenous suppressor of silencing. Furthermore, as CDKB1;1 interacted with DEL3, we propose that DEL3 is regulated by CDKB1;1 activity, consistent with a second expression peak of DEL3 at G2/M (Menges et al., 2005). Interestingly, the mitotic CDKB1;1, and not CDKA;1, co-purified with RBR,

providing further evidence that the E2F/DP/RB network is not only active at G1/S but also at G2/M transitions as was previously suggested in plants (Magyar et al., 2005) and as it is the case in *Drosophila* (Neufeld et al., 1998) and mammalian cells (Ishida et al., 2001).

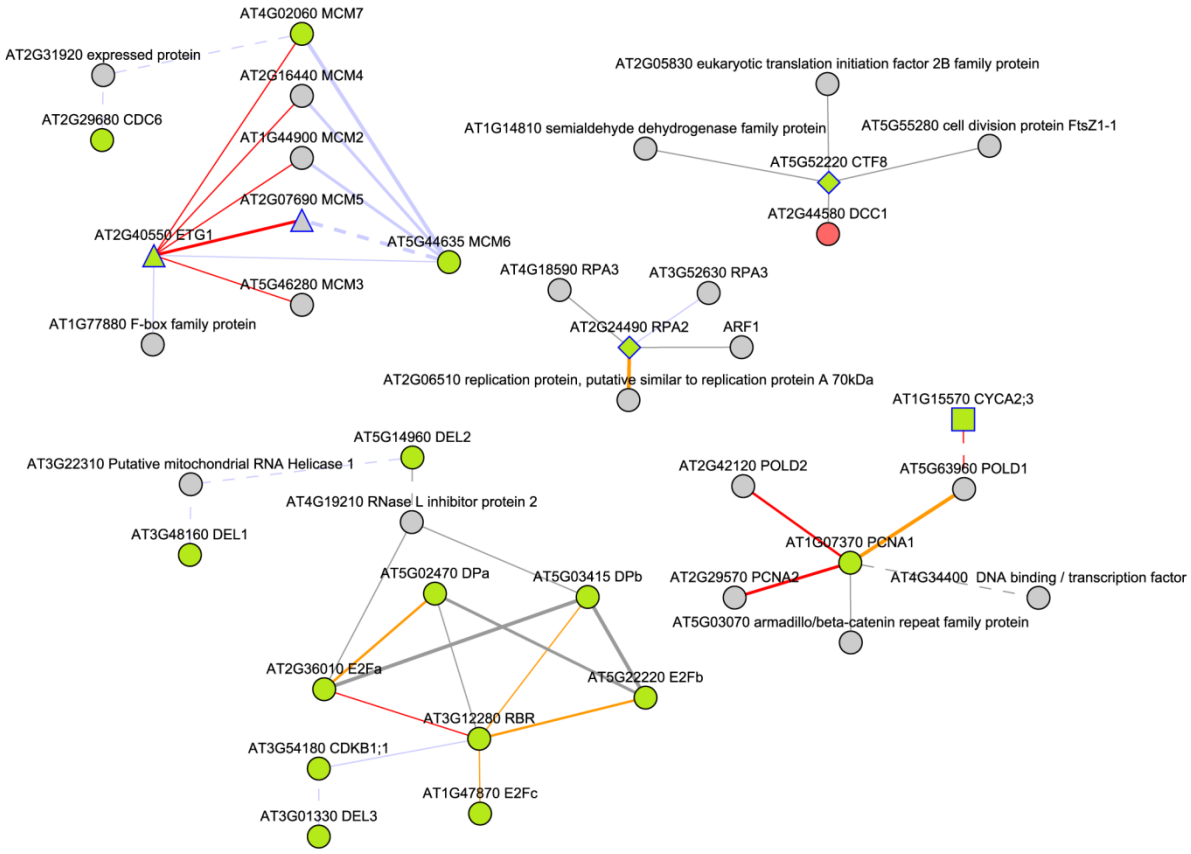


Figure 13. Representation of the E2F/DP/RB subnetwork and 4 other protein complexes involved in DNA replication

Once activated, E2F/DP complexes induce expression of genes required for DNA replication. DNA replication is initiated as the origin of replication complex (ORC), encoded by 7 genes in *Arabidopsis*, is established at the origins of replication (Shultz et al., 2007). TAP purification of ORC1a did not render any core interactors, possibly due to sterical hindrance of the TAP-tag. Next, Cdc6 binds the ORC and recruits the CDT/minichromosome maintenance complex (MCM2-7) involved in DNA replication licensing (Shultz et al., 2007). With Cdc6 we pulled down an expressed protein that also interacted with MCM7. The MCM complex, possessing helicase activity for unwinding of double stranded DNA, was isolated with MCM6 as bait, together with the recently published (Takahashi et al., 2008) and highly co-expressed E2F-

target gene 1 (ETG1). Once the pre-replication complex consisting of ORC/Cdc6/MCM2-7 is formed, it enables recruitment of the replication machinery (Shultz et al., 2007). The co-purified fraction of proliferating cell nuclear antigen 1 (PCNA1), a sliding clamp for DNA polymerase and thus a key actor in DNA replication, contained PCNA2, two DNA polymerase delta subunits (POLD1-2), of which one also interacted with CycA2;3, an armadillo/beta-catenin repeat family of unknown function and a DNA binding protein. Finally, we proof the existence of the alternative Ctf18 replication factor C complex in plants, required for sister chromatid cohesion in yeast (Mayer et al., 2001) and a protein complex involved in stabilization of single stranded DNA during replication, repair and transcription, including RPA2, 2 RPA3 proteins and a putative replication protein (AT2G06510) (Shultz et al., 2007).

The Anaphase Promoting Complex and mitotic checkpoint proteins

The anaphase promoting complex (APC) is an E3 ubiquitin ligase and targets cell cycle proteins for irreversible destruction by the 26S proteasome during mitosis and G1 (Capron et al., 2003a), giving direction to the cell cycle. Sequence homology revealed the existence of all APC core vertebrate counterparts in the *Arabidopsis* genome (Capron et al., 2003a). All APC core genes, with the exception of APC3/Cdc27, are unique and elimination can impair female gametophyte development, as was proven for APC6/Cdc16 (Kwee and Sundaresan, 2003) and APC2 (Capron et al., 2003b). In contrast to the APC core subunits, multiple APC activators exist in *Arabidopsis* (6 Cdc20 and 3 CCS52 (Cdh1-related) genes). As these APC activators also confer substrate specificity to the APC, the existence of 9 putative activators possibly reflects the need to target the many mitotic cyclins and other substrates present. Although Y2H has revealed interactions between different plant APC subunits, biochemical purification of the APC from plants has not been reported so far. Here we present for the first time the isolation of a plant APC (Fig. 14). The APC is visible as a very tightly interconnected subnetwork enriched for highly co-expressed gene pairs. We identified all putative plant APC subunits, except APC13 (Bonsai) and Cdc26, and show that both Cdc27a and Cdc27b/HOBBIT are part of the APC.

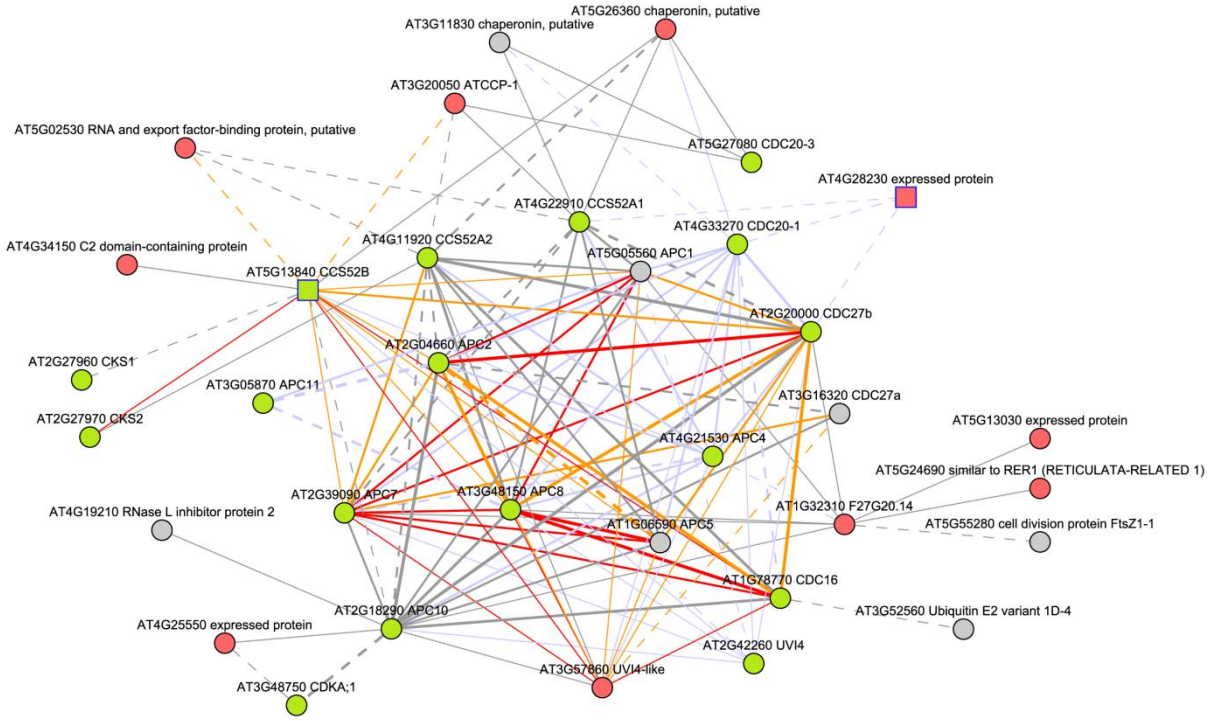


Figure 14. The anaphase promoting complex

In addition, we discovered three possible plant specific APC subunits: UVI4, UVI4-like and AT1G32310. UVI4 and UVI4-like co-purified with different APC subunits and activators and reverse purifications confirmed their interaction with the APC. When mutated, plants showed decreased sensitivity to UV-B stress due to an increase in endoreduplication (Hase et al., 2006). So here, we strengthen the previous reported statement that UVI4 functions in the maintenance of the mitotic state of the cell cycle, and link its function with the APC. Protein sequence analysis revealed different CDK consensus motifs in their sequence, possibly important in the regulation of their activity. Interestingly, both proteins have a C-terminal methionine-arginine tail. This MR-tail, present in only 40 *Arabidopsis* proteins, is in Xenopus involved in cdc20-independent binding of Nek2a to the APC (Hayes et al., 2006) and resembles the known isoleucine-arginine tail present in the APC activators, involved in binding of the APC activators to the tetratricopeptide repeat containing APC subunits (Vodermaier et al., 2003). Next to UVI4 and UVI4-like, an unknown protein (AT1G32310 F27G20.14) co-purified with the APC and this was confirmed by reverse TAP. Known APC substrates like mitotic cyclins were not identified, perhaps due to their unstable character and the long purification time. However we picked up CKS1 and CKS2, docking factors of

CDK-cyclin complexes and known APC substrates in yeast. CDKA;1 interacted with APC10, indicating regulation of APC activity by CDKA;1. Also in plants, APC activators are probably assembled on the APC by the action of the CCT Chaperonin (Camasses et al., 2003) as 3 members co-purified. Besides, this chaperonin could assist in the assembly of the spindle checkpoint complex as we show interaction with 3 mitotic checkpoint proteins (Fig. 15). This complex monitors progress during mitosis, delaying anaphase onset until each chromosome is correctly attached to the spindle. Unattached kinetochores trigger formation of a Mad2/Bub3/BubR1 complex, which in turns inhibits Cdc20 APC activators, thereby preventing degradation of several cell cycle regulators and progression of anaphase (Kimbara et al., 2004).

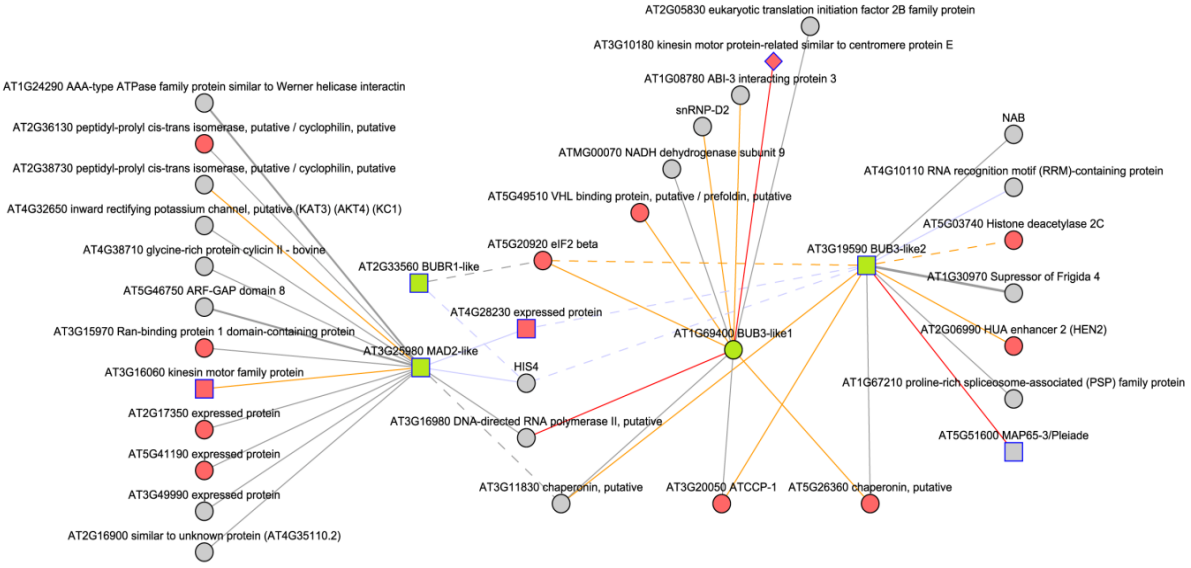


Figure 15. Network around 4 putative mitotic checkpoint proteins

We did not purify Cdc20, however the putative mitotic checkpoint proteins co-purified many specific interactors, including M-phase specific kinesins, of which one is similar to centromere protein E, the highly co-expressed MAP65-3, located at mitotic microtubule arrays and essential for cytokinesis (Muller et al., 2004), 2 peptidyl-prolyl cis-trans isomerases, 2 proteins of the prefoldin chaperone, a helicase (AT1G24290) similar to a Replication factor C protein and an ADP-ribosylation factor GTPase-activating protein (AT3G15970). The latter two proteins were also predicted to interact with MAD2-like (Geisler-Lee et al., 2007). Histone H4 was pulled down with 3 mitotic checkpoint proteins.

Based on interactions with microtubule associated proteins and proteins involved in DNA replication, we are probably dealing with *bona fide* spindle assembly checkpoint proteins. Additional interactors may hint for a function in regulation of transcription (DNA-directed RNA polymerase II), splicing (AT1G67210) and mRNA stability (HUA ENHANCER2).

4.3 Conclusions and perspectives

Knowledge of the basic cell cycle machinery is a prerequisite to grasp how signaling pathways impinge on it and regulate cell proliferation during plant growth and development in a changing environment. Genome annotation of plants, followed by sequence homology searches, revealed the existence of many genes involved in cell proliferation (Vandepoele et al., 2002; Capron et al., 2003a; Menges et al., 2005; Shultz et al., 2007). Microarray analysis showed that many of these have a cell cycle-dependent expression profile (Menges et al., 2005), sustaining their role in the regulation of this process. To add the next level of functional annotation, we conducted a targeted functional proteomics approach in *Arabidopsis thaliana*.

Protein complexes around cell cycle baits were purified using our TAP platform. After filtering, a robust and highly connected cell cycle network was achieved. This interactome map was enriched in highly co-expressed gene pairs, in genes belonging to the same GO category, and in genes containing cell cycle-related features. We confirmed previously known and predicted interactions, proving the fidelity of our interactome, and extended the cell cycle network, as a high percentage of the interactions were not reported before. Out of 6 reverse purifications, we could confirm the initial interaction every time, fortifying the predictive power of the interactome. Through a computational analysis, we provide a list of 40 candidate new cell cycle proteins from the core dataset, demonstrating that starting from a certain bait list, one can further extend the pathway of interest. Although computational analysis has proven the quality of the non-core dataset, it is still a combination of *bona fide* interactors and spurious background proteins. So, to further filter out these spurious proteins, one could screen for their frequency in a TAP interaction dataset obtained with baits not linked with cell cycle, as we have such a dataset in-house. Once filtered, additional information could be further integrated to select the most interesting interactors. For

instance, links between the cell cycle pathway and hormone signalling could be discovered by screening for preys containing hormone-responsive cis-regulatory elements in their promoter or for preys encoded by genes that are differentially expressed in hormone-related transcript expression analysis. Their interaction can then be confirmed through reverse TAP experiments or through Co-IP, and finally their role in cell cycle or plant growth and development should be assessed through overexpression or knock-out approaches. Combining our interactome data with cell cycle-related expression profiles gives insight in which CDK/cyclin complexes are active during cell division and when. We picked up some interesting potential substrates, however, mainly due to technical limitations and the transient nature of these interactions, a fraction of them could not be confirmed. Nevertheless, the non-core dataset can be used as a source of potential substrates. To obtain a more comprehensive view on the substrate specificity of these CDK/cyclin complexes, these purified entities should be used in combination with genome-wide protein microarrays (Popescu et al., 2007). Furthermore, we present 3 gene networks concerning positive and negative regulation of cell division, followed by complexes involved in activation of gene regulatory pathways essential for cell cycle transitions, by complexes acting in DNA replication and repair, and spindle checkpoint complexes. Finally, and for the first time in plants, we purified the anaphase-promoting complex and identified 3 plant-specific APC subunits.

Integrating this protein interaction data with gene regulatory, gene expression, and metabolite networks will be a first step in obtaining a systems biology view on the plant cell cycle. This robust plant cell culture-based TAP platform is further an excellent tool to study interactome network dynamics by comparing interactomes identified under differential conditions. The system allows synchronization what would enrich the cells in protein complexes at certain time points in the cell cycle. Through starvation approaches one could study cell cycle exit and cell cycle re-entry into more detail. The TAP platform could be used to unravel how differentiation cues alter interactome networks by e.g. addition of peptides to stimulate differentiation (Whitford et al., 2008). To transfer knowledge obtained from this plant cell cycle interactome to other higher eukaryotes, it would be interesting to identify the novel orthologous interactions (interologs). Since this is the first comprehensive cell

cycle interactome for higher eukaryotes mapped through complex isolation, this approach could shed light on cell cycle regulation and its link with other processes in organisms like humans to study diseases like cancer or in *Caenorhabditis elegans*, and *Drosophila melanogaster* to further unravel developmental processes.

4.4 Materials & Methods

4.4.1 Experimental procedures

Cloning of transgenes encoding tag fusions, transformation of plant cell suspension cultures, protein extract preparation and TAP were done as previously described (Van Leene et al., 2007) (Chapter 2). The adapted protocol used for purification of protein complexes incorporating GS-tagged bait is described elsewhere (Van Leene et al., 2008) (Chapter 3). For identification by mass spectrometry, minor adjustments were implemented compared to previous described protocols (Van Leene et al., 2007), as shown below.

Proteolysis and peptide isolation

After destaining, gel slabs were washed for 1 hour in H2O, polypeptide disulfide bridges were reduced for 40 min in 25 mL of 6,66 mM DTT in 50 mM NH_4HCO_3 and sequentially the thiol groups were alkylated for 30 min in 25 mL 55 mM IAM in 50 mM NH_4HCO_3. After washing the gel slabs 3 times with water, complete lanes from the protein gels were cut into slices, collected in microtiter plates and treated essentially as described before with minor modifications (Van Leene et al., 2007). Per microtiter plate well, dehydrated gel particles were rehydrated in 20 µL digest buffer containing 250 ng trypsin (MS Gold; Promega, Madison, WI), 50 mM NH_4HCO_3 and 10% CH_3CN (v/v) for 30 min at 4° C. After adding 10 µL of a buffer containing 50 mM NH_4HCO_3 and 10% CH_3CN (v/v), proteins were digested at 37° C for 3 hours. The resulting peptides were concentrated and desalted with microcolumn solid phase tips (PerfectPureTM C18 tip, 200 nL bed volume; Eppendorf, Hamburg, Germany) and eluted directly onto a MALDI target plate (Opti-TOFTM384 Well Insert; Applied Biosystems, Foster City, CA) using 1.2 µL of 50% CH_3CN: 0.1% CF_3COOH solution saturated with α-cyano-4-hydroxycinnamic acid and spiked with 20 fmole/µL Glu1 Fibrinopeptide B (Sigma Aldrich), 20 fmole/µL des-Pro2-Bradykinin (Sigma Aldrich), and 20 fmole/µL Adrenocorticotropic Hormone Fragment 18-39 human (Sigma Aldrich).

Acquisition of mass spectra

A MALDI tandem MS instrument (4700 and 4800 Proteomics Analyzer; Applied Biosystems) was used to acquire peptide mass fingerprints and subsequent 1 kV CID fragmentation spectra of selected peptides. Peptide mass spectra and peptide sequence spectra were obtained using the settings essentially as previously presented (Van Leene et al., 2007). Each MALDI plate was calibrated according to the manufacturers' specifications. All peptide mass fingerprinting (PMF) spectra were internally calibrated with three internal standards at m/z 963.516 (des-Pro2-Bradykinin), m/z 1570.677 (Glu1-Fibrinopeptide B), and m/z 2465,198 (Adrenocorticotropic Hormone Fragment 18-39) resulting in an average mass accuracy of 5 ppm ± 10 ppm for each analyzed peptide spot on the analyzed MALDI targets. Using the individual PMF spectra, up to sixteen peptides, exceeding a signal-to-noise ratio of 20 that passed through a mass exclusion filter were submitted to fragmentation analysis.

MS based protein homology identification

PMF spectra and the peptide sequence spectra of each sample were processed using the accompanied software suite (GPS Explorer 3.6, Applied Biosystems) with parameter settings essentially as previously described (Van Leene et al., 2007). Data search files were generated and submitted for protein homology identification by using a local database search engine (Mascot 2.1, Matrix Science). An in house non-redundant *Arabidopsis* protein database called SNAPS *Arabidopsis thaliana* version 0.4 (SNAPS = Simple Non-redundant Assembly of Protein Sequences, 77488 sequence entries, 30468560 residues; available at http://www.ptools.ua.ac.be/snaps) was compiled from nine public databases. Protein homology identifications of the top hit (first rank) with a relative score exceeding 95% probability were retained. Additional positive identifications (second rank and more) were retained when the score exceeded the 98% probability threshold. Because identifications were done with different versions of the SNAPS database (Van Leene et al., 2007), and with the goal to obtain more uniformity between the identifications, all identifications from the core and the non-core dataset were resubmitted to Mascot and identified with the protein sequence repertoire from the latest TAIR database (TAIR8.0). Furthermore, an additional restriction was implemented to reduce the number of false positive identifications, and as

so, identifications, for which more than 50 % of the corresponding peptides had a trypsin miss-cleavage, were discarded.

4.4.2 Bioinformatics

Analysis of over- and underrepresentation of GO terms was done with the BiNGO tool (Maere et al., 2005) in Cytoscape (Shannon et al., 2003). The hypergeometric test was chosen at a significance value of 0.05 with the Benjamini and Hochberg False Discovery Rate Correction for multiple testing. The *Arabidopsis* gene annotation file used in the analysis was downloaded from the gene ontology website on the 4th of October 2008.

P-values for the enrichment analysis were calculated with the hypergeometric cumulative distribution function of the Matlab 7.5 software.

To calculate the GO similarity scores, GO terms (biological process and cellular component annotation) were extracted from the Gene Ontology database (Ashburner et al., 2000) and annotations for *Arabidopsis* proteins were downloaded from TAIR (Rhee et al., 2003). For each protein pair, all GO terms of both proteins were compared to each other. For each pair of GO terms, the depth of the common ancestor of the terms, which is the shortest path of the common ancestor to the root (GO:0003673), is calculated. Subsequently, the maximum value of the calculated depths is taken as the GO similarity score for a certain protein pair. GO term assignments based on physical interactions (IPI) or electronically assigned and less reliably assigned GO terms (with evidence codes ND, NR, NAS and IEA) were removed.

Network graphs were build using the Cytoscape software (Shannon et al., 2003).

4.5 Supplementary data

Supplemental material is provided in the CD-ROM accompanying the thesis. This file contains the list with control identifications, a cell cycle collection of 518 genes annotated by GO or described to be involved in cell cycle, the results of the GO analysis from BiNGO, an overview of the microarray experiments used to calculate the PCCs, and the list of candidate new cell cycle genes extracted from the core or the non-core dataset. Further, a Cytoscape file is provided containing the whole filtered dataset, the core and non-core dataset and the discussed subnetworks, together with an Excel file representing the interactions in a matrix

model. Finally, we provide supplemental MS-related information used for the unambiguous protein identification. This data includes PMF and MSMS settings, the m/z exclusion list, Sodium and Potassium adduct exclusion settings, settings for spectral interpretation, the spectral peak exclusion list, the used search parameter settings and the supplemental protein identification table.

Acknowledgments

We thank all members of the Functional Proteomics group at the Plant Systems Biology Department and all members of the Centre for Proteomics and Mass spectrometry at Antwerp University for technical assistance. This work was supported by grants from the Institute for the Promotion of Innovation through Science and Technology in Flanders (Generisch Basisonderzoek aan de Universiteiten [GBOU] grant no. 20193) and from the European Commission (Agronomics; LSHG-CT-2006-037704).

Notes to Chapter 4:

Summary

A crucial step towards understanding cellular systems is mapping networks of physical protein-protein interactions. The main objective of this work was bipartite: first we wanted to develop a generic technology platform that allows mapping of protein-protein interactions from *Arabidopsis thaliana* through protein complex isolation and protein characterization by mass spectrometry. Second, this platform would be used to unravel the cell cycle interactome of *Arabidopsis thaliana*. As this plant represents the model organism of higher plants, this interactome would certainly increase our knowledge of plant growth and development, and of the basic cell cycle machinery of plants and other higher eukaryotes in particular.

Due to the high degree of structural and physicochemical heterogeneity of proteins, which is directly linked to the diversity of their functions within a cell, the analysis of proteins is much more technically demanding than genome or transcriptome analyses. However, recent technical advances sustaining the emergence of powerful methods as mass spectrometry allow a system-wide study of proteins. The research field that analyses proteins at a system-wide level is called proteomics and covers different aspects of proteins: a proteome study is expected to represents a comprehensive survey of all proteins expressed at a specific time point, under certain conditions, in a given tissue. Furthermore, in addition to their primary amino-acid sequence, other properties of proteins such as their association with other proteins or molecules of different types, their state of modification, their relative amounts, specific activity, subcellular localization, and three-dimensional structure, represent crucial information to understand the function of a protein and ultimately, for the description of complex biological systems. In **Chapter 1**, we give an overview of these proteomic technologies and exemplify them with relevant studies conducted in plant species. The main emphasis of this introductory chapter however, lies in the discussion of methods enabling analysis of protein-protein interactions, as this is the topic of the experimental work presented in this thesis. In addition, a state of the art overview is given of available comprehensive protein interactomes from other organisms like *Saccharomyces cerevisiae*.

To date, the most widely used methods to identify protein-protein interactions are on the one hand the Yeast Two-Hybrid (Y2H) approach, allowing mapping of binary transient and stable interactions, and on the other hand approaches based on isolation of native protein complexes through affinity chromatography and protein identification by mass spectrometry. As plant protein interactions are investigated in a heterologous system with the Y2H method, isolation of protein complexes resembles more an *in vivo* representation of an interactome. To facilitate high-throughput mapping through affinity chromatography, methods were developed targeting a protein of interest through protein tagging approaches. One-step purification methods were with success applied in some model organisms as yeast and human, however the low specificity of this approach generates high false positive rates. So a more elegant approach was developed in yeast, based on two consecutive affinity purification steps, called the tandem affinity purification (TAP) method (Rigaut et al., 1999). In TAP, a combination of two high-affinity handles is used to reduce background caused by spurious and promiscuous proteins, while maintaining a high protein complex purification yield.

As described in **Chapter 2**, we developed a TAP-based technology platform to isolate and characterize protein complexes from suspension-cultured *Arabidopsis thaliana* cells (Van Leene et al., 2007). This platform covers all steps of the TAP approach: multisite Gateway-based vectors were designed, allowing high-throughput and flexible cloning of expression vectors encoding TAP tagged proteins of interest. This system allows either C- or N-terminal cloning of a desired open-reading frame to any TAP tag. Moreover, transcription can be driven from a promoter of choice, like the endogenous promoter, the constitutive Tobacco Mosaic Virus 35S promoter or an inducible promoter, although we show that it is beneficial to use the 35S promoter to compete with the endogenous protein for complex assembly. A protocol was optimized for efficient transformation of these transgenes in the *Arabidopsis thaliana* cell suspension culture, named PSB-d. These fast dark-growing cell suspension cultures were chosen as it provides an ideal system to study protein complexes involved in cell division, in light of the second major deliverable of the project. Transformation is mediated through co-cultivation of *Agrobacterium tumefaciens* containing the transgene with these suspension-cultured plant cells. Originally, this co-cultivate was plated out on

selective solid plant medium and after callus growth, transformants were selected based on GFP expression and re-cultured in liquid medium. However, the transformation protocol was shortened by direct selection of transformed cells in liquid medium. Next, methods were developed for fast up-scaling of these cultures, and for efficient total protein extract preparation from these plant cells. The yeast tandem affinity purification protocol was adapted for isolation of cell cycle protein complexes from these plant protein extracts. In collaboration with the Centre for Proteomics and Mass spectrometry of the University of Antwerp, an efficient work flow was set up allowing unambiguous protein identification of one-dimensional separated and trypsinized proteins by matrix-assisted laser desorption/ionization tandem mass spectrometry. As proof of concept, activity of purified complexes is demonstrated by a functional assay and a first protein interaction network of 42 reproducible interactions obtained with 6 cell cycle bait proteins is presented.

Despite the successful transformation of the yeast TAP method to *Arabidopsis*, some problems were still associated with the technology when applying plant protein extracts, as illustrated by the low number of TAP purified plant complexes reported so far. As in detail discussed in **Chapter 3,** the traditional TAP method still faces shortcomings as low specificity and low protein complex yield when applied in higher eukaryotes. So to further optimize the method and to bring protein complex analysis from plants tissues to its full bloom, we were continuously in search for optimized versions of the TAP tag, as it is the purification itself that at that stage would make the difference. So far, we evaluated 6 different TAP tags of which 4 were designed in house. Improved results concerning purification specificity and complex yield as compared to the traditional TAP tag were obtained with two different TAP tags, namely the GS tag, developed for TAP in mammalian cells (Burckstummer et al., 2006), and the in-house developed CSFH tag, providing alternatives when one of the two tags fails (Van Leene et al., 2008). Although results obtained with the four other alternative TAP tags were not satisfying, this analysis provided us with the necessary insight for future TAP tag design, and 4 alternatives are currently under evaluation, designed for higher recovery of transient interactions. This alternative TAP tag evaluation screen is discussed into more detail in Chapter 3 and in the accompanying **supplement**.

Finally, as discussed in **Chapter 4**, this TAP-based technology platform was used to map the cell cycle interactome of *Arabidopsis thaliana*. For this purpose, an in house available cell cycle ORFeome was extended and fused to the traditional TAP tag or the improved GS tag. The previously described technology platform was subsequently used to study protein complexes among 108 proteins involved in cell cycle. These proteins comprise key players governing progression through the cell cycle, like cyclin-dependent kinases (CDKs), their cyclin partners, additional regulatory proteins of these CDK/cyclin complexes, proteins of the E2F/DP/RB pathway, proteins of the anaphase promoting complex (APC), mitotic checkpoint proteins, and proteins of the core DNA replication machinery. Most of these cell cycle regulators were identified through sequence homology searches (Vandepoele et al., 2002; Capron et al., 2003a; Menges et al., 2005; Shultz et al., 2007), and microarray analysis and mutant analysis confirmed their role in cell division. Of these 108 tested baits, 102 proteins were successfully expressed as TAP fusions and subjected to TAP. This generated a high-confidence core dataset of 371 experimentally confirmed interactions among 196 proteins and a non-core dataset of 486 interactions among 320 proteins. To test the novelty of the interactome, different protein-protein interaction databases were screened, revealing that 66 % of the core and 95 % of the non-core dataset was not reported before. The quality of the cell cycle interactome was assessed through an integrative approach combining transcript co-expression values, gene ontology similarities, and cell cycle-related features. This analysis revealed an enrichment in both datasets for genes containing an E2F or mitosis-specific activator (MSA) motif in their promoter sequence. The core dataset was further enriched for genes that are periodically expressed during the cell cycle. On the other hand, the non-core dataset was enriched for proteins containing a CDK phosphorylation site, sustaining our hypothesis that this dataset is biased towards transient interactions linking the cell cycle machinery with other pathways. Integration of transcript co-expression values revealed that a high fraction of the gene pairs in our interactome are highly co-expressed. A list of 40 candidate new cell cycle genes is provided based on a integration of cell cycle features. Biological important gene networks were extracted from the interactome and are discussed. We present different CDK/cyclins complexes, speculate about their time of action through integration with cell cycle specific transcript information, and propose different

interesting substrates of these CDK/cyclin complexes. A new link between the CDK-activating kinase CDKF;1 and G-type CDKs is discovered. We found CycL;1 as the cyclin partner of these G-type CDKs, and interactors hint for a function of these CDKG/CycL;1 complexes in regulation of splicing and transcription. Two gene networks are further shown involved in negative regulation of CDK/cyclin complexes, and we state that mitotic CDK/cyclin complexes are directly regulated through interaction with Siamese and Siamese-related proteins (SMR1-2). We extend the function of the E2F/DP/RB pathway from the G1/S transition to the G2/M transition. For the first time ever, we were able to isolate the plant APC, and we identified 3 new APC subunits. Finally, we demonstrate the existence of different complexes involved in DNA synthesis and DNA repair in plants, and discovered a whole new unknown network around the mitotic spindle checkpoint proteins. In conclusion, this cell cycle interactome may serve as a hypothesis-generating tool to further extend our knowledge of cell division and plant growth and development.

Samenvatting

Een cruciale stap in het begrijpen van cellulaire systemen is het in kaart brengen van netwerken die fysische eiwit-eiwit interacties weergeven. De hoofddoelstelling van dit werk was tweeledig: eerst en vooral werd er getracht om een generisch technologie platform te ontwikkelen waarmee eiwit-eiwit interacties in kaart kunnen gebracht worden aan de hand van isolatie van eiwitcomplexen en eiwit karakterisering door massaspectrometrie. Ten tweede, zou dit platform gebruikt worden om het cel cyclus interactoom van *Arabidopsis thaliana* te ontrafelen. Daar deze plant het model organisme vertegenwoordigt van hogere planten, zal dit interactoom ongetwijfeld onze kennis inzake plantengroei en ontwikkeling, en inzake de celdeling machinerie van planten en andere hogere eukaryoten in het bijzonder, doen toenemen.

Wegens de hoge graad van structurele en fysico-chemische heterogeniteit van eiwitten, dewelke onmiddellijk kan gekoppeld worden aan de diversiteit van hun functies binnen een cel, vereist de analyse van eiwitten veel meer technisch vernuft dan genoom of transcriptoom analyses. Desondanks hebben recente technische vooruitgangen, die de opkomst van krachtige methodes als massaspectrometrie ondersteunen, er voor gezorgd dat eiwitten kunnen bestudeerd worden op een grote schaal. Het onderzoeksgebied dat eiwitten op systeemwijde schaal analyseert wordt ‘proteomics’ genoemd en omvat verschillende aspecten van eiwitten: zo wordt er van een proteoom-analyse verwacht dat deze een uitgebreid beeld weergeeft van alle eiwitten die tot expressie komen op een bepaald ogenblik, onder bepaalde condities in een gegeven weefsel. Ook andere eigenschappen van eiwitten, naast hun primaire aminozuursequentie, zoals met welke andere eiwitten of biomoleculen ze associëren, hun staat van modificatie, hun relatieve hoeveelheid, specifieke activiteit, subcellulaire lokalisering, en 3-dimensionele structuur, leveren belangrijke informatie over de functie van een eiwit en uiteindelijk voor de beschrijving van complexe biologische systemen. In **Hoofdstuk 1** geven we een overzicht van deze proteoom technieken en verduidelijken we alles aan de hand van relevante voorbeelden uit het planten onderzoeksgebied. De nadruk van dit inleidend hoofdstuk ligt echter in de bespreking van de methoden die de analyse van eiwit-eiwit interacties toelaten, daar dit het hoofdonderwerp is van het experimentele werk dat werd uitgevoerd in dit

proefschrift. Bovendien wordt er hier een overzicht gegeven van eiwit interactomen van andere organismen als bakkersgist die reeds voor handen zijn.

Vandaag de dag berusten de meest gebruikte methoden om eiwit-eiwit interacties te identificeren op het gist twee-hybride systeem dat het in kaart brengen van binaire, zowel kortstondige als stabiele, interacties toelaat, en op methoden die gebaseerd zijn op de isolatie van endogene eiwit complexen aan de hand van affiniteitchromatografie en eiwit identificatie met behulp van massaspectrometrie. Daar planten eiwit interacties met de gist twee-hybride technologie bestudeerd worden in een heteroloog systeem, geeft de isolatie van eiwit complexen een meer representatief beeld van een *in vivo* interactoom. Om hoge doorvoer van eiwitcomplex zuivering toe te laten werden er methodes ontwikkeld waarbij het eiwit van interesse bestudeerd wordt via eiwit 'tagging'. Zuiveringen gebaseerd op 1 verrijkingsstap werden met succes toegepast in gist en humane cellen, maar door de lage specificiteit van deze methode wordt een hoge fractie vals positieve interacties bekomen. Daarom werd in gist een meer elegante toepassing ontwikkeld die gebaseerd is op twee opeenvolgende zuiveringsstappen, tandem affiniteit zuivering (TAP) genoemd (Rigaut et al., 1999). Bij TAP wordt er gebruik gemaakt van een combinatie van twee hoge affiniteitgrepen om de achtergrond die veroorzaakt wordt door o.a. hoog abundante eiwitten te verminderen en tegelijk de opbrengst van de gezuiverde complex hoog te houden.

Zoals het beschreven is in **Hoofdstuk 2**, hebben we een technologie platform ontwikkeld gebaseerd op TAP om eiwit complexen uit in suspensie gebrachte plantencellen van *Arabidopsis thaliana* te isoleren en te karakteriseren (Van Leene et al., 2007). Dit platform omvat alle stappen betrokken in de TAP methode: er werden vectoren gemaakt gebaseerd op het 'multisite-Gateway' systeem. Dit laat een flexibele en hoge doorvoer van klonering van expressievectoren, coderend voor een fusie van het eiwit van interesse aan de TAP-tag, toe. Met dit systeem is het mogelijk om gelijk welke TAP-tag aan ofwel het C-terminale uiteinde ofwel het N-terminale uiteinde van een gewenst open leesraam te fusioneren. Bovendien kan de transcriptie van het 'bait'-coderend transgen gedreven worden door een promotor van keuze, zij het de endogene promotor, de 35S constitutieve promotor van het tabaksmozaïek virus, of een induceerbare promotor. Nochtans hebben we aangetoond dat

het voordelig is om een constitutieve promotor te gebruiken om competitie met het endogene eiwit voor assemblage in eiwit complexen te verhogen. Er werd vervolgens een protocol op punt gesteld voor een efficiënte transformatie van deze transgenen in de *Arabidopsis thaliana* suspensie cultuur genaamd PSB-d. Deze snel en in het donker groeiende cultuur werd gekozen omdat ze een ideaal systeem aanbiedt voor de studie van eiwit complexen betrokken in celdeling, dit met het oog op het behalen van de tweede hoofddoelstelling. Transformatie wordt uitgevoerd via co-cultivatie van *Agrobacterium tumefaciens* cellen die het transgen van interesse bevat met de in suspensie gebrachte plantencellen. Oorspronkelijk werd dit co-cultivaat uitgeplaat op een selectieve vaste plantengroeibodem, en na groei van het callus weefsel werden getransformeerde cellen geselecteerd op basis van de GFP-expressie en opnieuw in vloeibaar medium gebracht. Dit transformatie protocol werd ingekort door directe selectie van getransformeerde cellen in vloeibaar medium. Vervolgens werden er methoden ontwikkeld voor de vlugge opschaling van deze culturen en voor een efficiënte extractie van totaal eiwit uit deze plantencellen. Het gist tandem affiniteit zuiveringsprotocol werd vervolgens aangepast voor de isolatie van eiwitcomplexen betrokken in celdeling uit planten cel extracten. In samenwerking met Ceproma, het centrum voor proteoom analyse en massaspectrometrie van de universiteit van Antwerpen, werd er een efficiënte pijplijn opgesteld die eiwit identificatie toelaat van 1-dimensioneel gescheiden en met trypsine in peptiden geknipte eiwitten via MALDI/MSMS. Als finaal bewijs van de werking van dit platform, hebben we activiteit aangetoond van gezuiverde complexen in een functionele analyse en een eerste eiwit-eiwit interactie netwerk van 42 experimenteel herhaalbare interacties bekomen met 6 eiwitten betrokken in celdeling werd in kaart gebracht.

Ondanks de succesvolle overdracht van de TAP methode van gist naar *Arabidopsis*, waren er nog enkele problemen verbonden aan de technologie bij het gebruik van plantenextract uit planten, getuige het laag aantal rapporteringen van complexen gezuiverd via TAP in planten. Zoals in detail besproken wordt in **Hoofdstuk 3**, gaat de traditionele TAP methode gepaard met enkele tekortkomingen zoals lage specificiteit en een lage opbrengst van gezuiverde complexen wanneer toegepast in hogere eukaryoten. Dus om de methode verder te optimaliseren en om eiwit complex analyse uit plantenweefsel tot zijn volle recht te laten

komen, waren we continu op zoek naar verbeterde versies van de TAP tag, daar het de zuivering zelf was die op dat moment het verschil kon maken. Tot nu toe hebben we 6 verschillende TAP tags geëvalueerd waarvan er 4 binnenshuis ontworpen zijn. Betere resultaten qua zuiverheidgraad en complexopbrengst werden bekomen met twee verschillende TAP tags, namelijk met de GS tag die ontworpen is voor gebruik in zoogdiercellen (Burckstummer et al., 2006) en de zelf ontworpen CSFH tag, dat zo een alternatief biedt voor wanneer de GS tag faalt (Van Leene et al., 2008). Ondanks dat de resultaten met de andere alternatieve TAP tags teleurstellend waren, heeft deze analyse ons het nodige inzicht bijgebracht voor toekomstige TAP tag ontwerpen, en momenteel worden er 4 nieuwe TAP tags uitgetest met het oog op het stabiliseren van kortstondige interacties. Deze evaluatie van alternatieve TAP tags komt uitvoerig aan bod in hoofdstuk 3 en het bijgevoegd **supplement**.

Zoals tenslotte in **Hoofdstuk 4** besproken wordt, werd dit TAP-gebaseerd platform gebruikt om het celdelings-interactoom van *Arabidopsis thaliana* in kaart te brengen. Om deze doelstelling te bereiken, werd een bestaande collectie van open leesramen gerelateerd aan celdeling uitgebreid en gefusioneerd aan de traditionele TAP tag of aan de verbeterde GS versie. Het voorgaand beschreven technologie platform werd vervolgens gebruikt om eiwit complexen rond 108 eiwitten betrokken in celdeling te bestuderen. Deze eiwitten omvatten de sleutelpionnen die het verloop van de celdeling reguleren, zoals cycline-afhankelijke kinases (CDKs), hun cycline partners, additionele regulerende eiwitten van deze CDK/cycline complexen, eiwitten uit het E2F/DP/RB netwerk, eiwitten van het Anafase-Bevorderend-Complex (APC), mitotische controle eiwitten, en eiwitten behorende tot de kern van de DNA replicatie machinerie. Het grotendeel van deze regulerende eiwitten werd geïdentificeerd via sequentie homologie speurtochten (Vandepoele et al., 2002; Capron et al., 2003a; Menges et al., 2005; Shultz et al., 2007), en via transcript en mutant analyse werd hun rol in celdeling bevestigd. Van deze 108 geteste doelwiteiwitten, kwamen er 102 succesvol tot expressie als TAP-fusie en deze werden onderworpen aan TAP. Dit gaf aanleiding tot een kern dataset van 371 hoog betrouwbare interacties die experimenteel bevestigd werden tussen 196 eiwitten en een niet-kern dataset van 486 interacties tussen 320 eiwitten. Om de vernieuwende waarde van het interactoom in te schatten, werd de overlap gemaakt met

data uit verschillende databanken, en deze analyse bracht aan het licht dat 66 % van de kern en 95 % van de niet-kern dataset voordien nog niet beschreven was. De kwaliteit van dit celdelings-interactoom werd gecontroleerd aan de hand van een aanpak die transcript co-expressie waarden, gen ontologie overeenkomsten en celdeling gerelateerde gen- en eiwitkenmerken integreert. Deze analyse bracht aan het licht dat beide datasets verrijkt zijn met genen die een E2F of mitotisch specifiek activator (MSA) motief in hun promotor sequentie bezitten. Bovendien was de kern dataset verder verrijkt voor genen die een fluctuerend transcript profiel vertonen tijdens celdeling. De niet-kern dataset daarentegen was verrijkt met eiwitten die een CDK fosforylatie motief bevatten, wat onze hypothese ondersteunt dat deze dataset een tendens heeft om kortstondige interacties weer te geven die de celdeling machinerie linken aan andere biologische processen. Integratie van transcript co-expressie waarden bewees dat een grote fractie van de gen paren meer dan willekeurig samen tot expressie komen. Een lijst van een 40-tal kandidaat nieuwe genen betrokken in celdeling werd voorspeld op basis van integratie van celdeling kenmerken. Biologisch belangrijke gen netwerken werden uit het interactoom geëxtraheerd en besproken. We stellen verschillende CDK/cycline complexen voor, speculeren over hun tijdstip van actie door combinatie met celdelings specifieke transcript informatie, en we identificeerden verschillende potentiële substraten van CDK/cycline complexen. Een nieuwe link werd gevonden tussen het CDK-activerend kinase CDKF;1 en G-type CDKs. We hebben bovendien ontdekt dat CycL;1 de cycline partner is van deze G-type CDKs en hun interactoren wijzen op een rol in regulatie van splicing en transcriptie. Verder worden er 2 netwerken aangebracht die betrokken zijn in de negatieve regulatie van CDK/cycline complexen, en beweren we dat mitotische CDK/cycline complexen rechtsreeks geregeld worden door interactie met Siamese of Siamese-gerelateerde eiwitten (SMR1-2). We breiden het werkterrein van het E2F/DP/RB netwerk uit van het G1/S transitiepunt naar de G2/M overgang. Bovendien hebben we voor de allereerste keer het planten APC geïsoleerd waarbij 3 nieuwe APC subeenheden ontdekt werden. Verder tonen we het bestaan aan van verschillende complexen betrokken in DNA synthese en herstel, en werd er een volledig nieuw netwerk bloot gelegd rond de mitotische spoel controle eiwitten. Tot slot kunnen we

besluiten dat dit interactoom dienst kan doen als een hypothese-genererend instrument om onze kennis omtrent celdeling en plantengroei en ontwikkeling uit te breiden.

Additional publications

A technology platform for the fast production of monoclonal recombinant antibodies against plant proteins and peptides.

Dominique Eeckhout, Annelies De Clercq, Eveline Van De Slijke, Jelle Van Leene, Hilde Stals, Peter Casteels, Geert Persiau, Dominique Vercammen, Frank Van Breusegem, Marc Zabeau, Dirk Inzé, Laurent Jespers, Ann Depicker, Geert De Jaeger.

Abstract

The application of recombinant antibodies in plant biology research is limited because plant researchers have minimal access to high-quality phage display libraries. Therefore, we constructed a library of 1.3 x 10(10) clones displaying human single-chain variable fragments (scFvs) that is available to the academic community. The scFvs selected from the library against a diverse set of plant proteins showed moderate to high antigen-binding affinity together with high specificity. Moreover, to optimize an scFv as immunodetection agent, two expression systems that allow efficient production and purification of bivalent scFv-Fc and scFv-CkappaZIP fusion proteins were integrated. We are convinced that this antibody platform will further stimulate applications of recombinant antibodies such as the diagnostic detection or immunomodulation of specific antigens in plants.

Manuscript published in Journal of Immunological Methods 294, 181-187 (2004)

A bio-analytical method for the proteome wide display and analysis of protein complexes from whole plant cell lysates.

Noor Remmerie, Luc Roef, Eveline Van De Slijke, Jelle Van Leene, Geert Persiau, Dominique Eeckhout, Hilde Stals, Kris Laukens, Filip Lemière, Eddy Esmans, Harry Van Onckelen, Dirk Inzé, Geert De Jaeger, Erwin Witters

Abstract

While protein interaction studies and protein network modeling come to the forefront, the isolation and identification of protein complexes in a cellular context remains a major challenge for plant science. To this end, a non-denaturing extraction procedure was optimized for plant whole cell matrices and the combined use of gel filtration and BN-PAGE for the separation of protein complexes was studied. Hyphenation to denaturing electrophoresis and mass spectrometric analysis allows for the simultaneous identification of multiple (previously unidentified) protein interactions in single samples. The reliability and efficacy of the technique was confirmed (I) by the identification of well-studied plant protein complexes, (II) by the presence of non-plant interologs for several of the novel complexes (III) by presenting physical evidence of previously hypothetical plant protein interactions and (IV) by the confirmation of found interactions using co-IP. Furthermore practical issues concerning the use of this 2-D BN/SDS-PAGE display method for the analysis of protein-protein interactions are discussed.

Manuscript published in Proteomics (2008), *in press*

The cyclin-dependent kinase inhibitor KRP6 impairs mitosis and cytokinesis through the inhibition of D-type cyclin/CDKA;1 complexes

Annelies De Clercq, Hilde Stals, Jelle Van Leene, Eveline Van De Slijke, Gert Van Isterdael, Dominique Eeckhout, Geert Persiau, Daniël Van Damme, Aurine Verkest, Anne Pharazyn, Erwin Witters, Harry Van Onckelen, Dirk Inzé, Lieven De Veylder, and Geert De Jaeger

Abstract

In *Arabidopsis thaliana*, seven cyclin-dependent kinase (CDK) inhibitors have been identified, designated interactors of CDKs (ICKs) or Kip-related proteins (KRPs). Here, the function of KRP6 was analyzed during cell cycle progression. First, the *in vivo* interaction partners of KRP6 were purified via tandem affinity purification from extracts of cultured *Arabidopsis* cells that over-express TAP-tagged KRP6. Mass spectrometry based protein identification identified CDKA;1 and two D-type cyclins, CYCD2;1 and CYCD3;1, as the *in vivo* binding partners of KRP6. *In vitro* kinase assays further demonstrated the inhibition of the activity of these two D-type cyclin/CDKA;1 complexes by KRP6. Surprisingly, *KRP6*-overexpressing suspension cultures displayed an accelerated entry into mitosis that coincided with an earlier increase of CDK activity. Although cells entered mitosis earlier, progression through and exit from mitosis was delayed. Phenotypic analysis revealed that this delayed progression correlated with the appearance of multinucleated cells that underwent defective mitosis and cytokinesis. These findings and the observed increase of KRP6 abundance upon treatment with a microtubuli depolymerizing drug strongly suggests that KRP6 acts as part of the spindle assembly checkpoint through the inhibition of D-type cyclin/CDKA;1 complexes.

Manuscript submitted to Plant Journal

Gene expression trends, protein features and GO annotations complement each other in an effective approach to gene function prediction

Krzysztof Wabnik, Torgeir Hvidsten, Anna Kedzierska, Thomas Skøt Jensen, Jelle Van Leene, Geert De Jaeger, Gerrit Beemster, Jan Komorowski and Martin Kuiper

Abstract

Motivation: Genome-scale 'omics' data are a sovereign source of biological information for data-driven systems biology approaches. The diversity and complexity of such information became the stumbling block to multi-data integration, and its correct interpretation. We proposed a novel data integration strategy to handle this particular issue. The gene expression data, protein features, and GO annotations were integrated in common patterns of biologically relevant information (If-Then rule models), and were used to predict the function of unknown genes in *Arabidopsis thaliana* and *Schizosaccharomyces pombe*.

Results: If-Then rule models were statistically validated on the training set showing success rates of up to 0.89 (discriminative and predictive power for both modelled organisms). Alternately, models built solely of one data type (protein features or microarray data) had success rates varying from 0.68 to 0.78. Our models were applied to generate classifications for unknown genes, of which a sizeable number were corroborated either by literature reports, or by computational approaches. Finally, we found a strong experimental evidence for these predictions compared to results from tandem affinity purification (TAP) experiments and by *in silico* experiments on the BioGRID interactome database – the significant fractions of predicted cell cycle interactors were confirmed by recent experimental study on cell cycle protein complexes. We demonstrated that gene expression data, protein features and GO annotations can be combined in one powerful approach to gene function prediction.

Manuscript published in Bioinformatics, *in press*

References

Abe, M., Fujiwara, M., Kurotani, K., Yokoi, S., and Shimamoto, K. (2008). Identification of dynamin as an interactor of rice GIGANTEA by tandem affinity purification (TAP). Plant Cell Physiol **49,** 420-432.

Adam, G.C., Sorensen, E.J., and Cravatt, B.F. (2002). Chemical strategies for functional proteomics. Mol Cell Proteomics **1,** 781-790.

Aebersold, R., and Goodlett, D.R. (2001). Mass spectrometry in proteomics. Chemical reviews **101,** 269-295.

Agrawal, G.K., Yonekura, M., Iwahashi, Y., Iwahashi, H., and Rakwal, R. (2005). System, trends and perspectives of proteomics in dicot plants Part I: Technologies in proteome establishment. J Chromatogr B **815,** 109-123.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell **92,** 291-294.

Anderson, L., and Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. Electrophoresis **18,** 533-537.

Archambault, V., Chang, E.J., Drapkin, B.J., Cross, F.R., Chait, B.T., and Rout, M.P. (2004). Targeted proteomic study of the cyclin-Cdk module. Molecular cell **14,** 699-711.

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006). Large-scale identification of protein-protein interaction of Escherichia coli K-12. Genome Res **16,** 686-691.

Arnau, J., Lauritzen, C., Petersen, G.E., and Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif **48,** 1-13.

Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J., and Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. Molecular and cellular biology **17,** 3094-3102.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics **25,** 25-29.

Backstrom, S., Elfving, N., Nilsson, R., Wingsle, G., and Bjorklund, S. (2007). Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. Molecular cell **26,** 717-729.

Baerenfaller, K., Grossmann, J., Grobei, M.A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., and Baginsky, S. (2008). Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science **320,** 938-941.

Baginsky, S., and Gruissem, W. (2006). Arabidopsis thaliana proteomics: from proteome to genome. Journal of experimental botany **57,** 1485-1491.

Barrios-Rodiles, M., Brown, K.R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R.S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I.W., Luga, V., Przulj, N., Robinson, M., Suzuki, H., Hayashizaki, Y., Jurisica, I., and Wrana, J.L. (2005). High-throughput mapping of a dynamic signaling network in mammalian cells. Science **307,** 1621-1625.

Batelli, G., Verslues, P.E., Agius, F., Qiu, Q., Fujii, H., Pan, S., Schumaker, K.S., Grillo, S., and Zhu, J.K. (2007). SOS2 promotes salt tolerance in part by interacting with the vacuolar H+-ATPase and upregulating its transport activity. Mol Cell Biol **27,** 7781-7790.

Bauer, A., and Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. European journal of biochemistry / FEBS **270,** 570-578.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villen, J., Li, J., Cohn, M.A., Cantley, L.C., and Gygi, S.P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A **101,** 12130-12135.

Benhar, I. (2007). Design of synthetic antibody libraries. Expert Opin Biol Ther **7,** 763-779.

Berggard, T., Linse, S., and James, P. (2007). Methods for the detection and analysis of protein-protein interactions. Proteomics **7,** 2833-2842.

Beyer, A., Bandyopadhyay, S., and Ideker, T. (2007). Integrating physical and genetic maps: from genomes to interaction networks. Nature reviews **8,** 699-710.

Bhat, R.A., Lahaye, T., and Panstruga, R. (2006). The visible touch: in planta visualization of protein-protein interactions by fluorophore-based methods. Plant Methods **2,** 12.

Bindschedler, L.V., Palmblad, M., and Cramer, R. (2008). Hydroponic isotope labelling of entire plants (HILEP) for quantitative plant proteomics; an oxidative stress case study. Phytochemistry **69,** 1962-1972.

Boniotti, M.B., and Gutierrez, C. (2001). A cell-cycle-regulated kinase activity phosphorylates plant retinoblastoma protein and contains, in Arabidopsis, a CDKA/cyclin D complex. Plant J **28,** 341-350.

Boudolf, V., Rombauts, S., Naudts, M., Inze, D., and De Veylder, L. (2001). Identification of novel cyclin-dependent kinases interacting with the CKS1 protein of Arabidopsis. Journal of experimental botany **52,** 1381-1382.

Boudolf, V., Barroco, R., Engler Jde, A., Verkest, A., Beeckman, T., Naudts, M., Inze, D., and De Veylder, L. (2004a). B1-type cyclin-dependent kinases are essential for the formation of stomatal complexes in Arabidopsis thaliana. The Plant cell **16,** 945-955.

Boudolf, V., Vlieghe, K., Beemster, G.T., Magyar, Z., Torres Acosta, J.A., Maes, S., Van Der Schueren, E., Inze, D., and De Veylder, L. (2004b). The plant-specific cyclin-dependent kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreduplicating cells in Arabidopsis. The Plant cell **16,** 2683-2692.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A.M., Schirle, M., Schlegl, J., Schwab, M., Stein, M.A., Bauer, A., Casari, G., Drewes, G., Gavin, A.C., Jackson, D.B., Joberty, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol **6,** 97-105.

Bracha-Drori, K., Shichrur, K., Katz, A., Oliva, M., Angelovici, R., Yalovsky, S., and Ohad, N. (2004). Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. Plant J **40,** 419-427.

Brajenovic, M., Joberty, G., Kuster, B., Bouwmeester, T., and Drewes, G. (2004). Comprehensive proteomic analysis of human Par protein complexes reveals an interconnected protein network. J Biol Chem **279,** 12804-12811.

Broder, Y.C., Katz, S., and Aronheim, A. (1998). The ras recruitment system, a novel approach to the study of protein-protein interactions. Curr Biol **8,** 1121-1124.

Brukhin, V., Gheyselinck, J., Gagliardini, V., Genschik, P., and Grossniklaus, U. (2005). The RPN1 subunit of the 26S proteasome in Arabidopsis is essential for embryogenesis. The Plant cell **17,** 2723-2737.

Bunnell, B.A., Heath, L.S., Adams, D.E., Lahti, J.M., and Kidd, V.J. (1990). Increased expression of a 58-kDa protein kinase leads to changes in the CHO cell cycle. Proc Natl Acad Sci U S A **87,** 7467-7471.

Burckstummer, T., Bennett, K.L., Preradovic, A., Schutze, G., Hantschel, O., Superti-Furga, G., and Bauch, A. (2006). An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat Methods **3,** 1013-1019.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature **433,** 531-537.

Camasses, A., Bogdanova, A., Shevchenko, A., and Zachariae, W. (2003). The CCT chaperonin promotes activation of the anaphase-promoting complex through the generation of functional Cdc20. Mol Cell **12,** 87-100.

Canovas, F.M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H.P., and Rossignol, M. (2004). Plant proteome analysis. Proteomics **4,** 285-298.

Capron, A., Okresz, L., and Genschik, P. (2003a). First glance at the plant APC/C, a highly conserved ubiquitin-protein ligase. Trends in plant science **8,** 83-89.

Capron, A., Serralbo, O., Fulop, K., Frugier, F., Parmentier, Y., Dong, A., Lecureuil, A., Guerche, P., Kondorosi, E., Scheres, B., and Genschik, P. (2003b). The Arabidopsis anaphase-promoting complex or cyclosome: molecular and genetic characterization of the APC2 subunit. The Plant cell **15,** 2370-2382.

Casati, P., Zhang, X., Burlingame, A.L., and Walbot, V. (2005). Analysis of leaf proteome after UV-B irradiation in maize lines differing in sensitivity. Mol Cell Proteomics **4,** 1673-1685.

Castro, A., Vigneron, S., Bernis, C., Labbe, J.C., and Lorca, T. (2003). Xkid is degraded in a D-box, KEN-box, and A-box-independent pathway. Mol Cell Biol **23,** 4126-4138.

Causier, B., and Davies, B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. Plant molecular biology **50,** 855-870.

Chang, I.F. (2006). Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. Proteomics **6,** 6158-6166.

Chang, I.F., Szick-Miranda, K., Pan, S., and Bailey-Serres, J. (2005). Proteomic characterization of evolutionarily conserved and variable proteins of Arabidopsis cytosolic ribosomes. Plant physiology **137,** 848-862.

Cheeseman, I.M., and Desai, A. (2005). A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. Sci STKE **2005,** pl1.

Cheeseman, I.M., Brew, C., Wolyniak, M., Desai, A., Anderson, S., Muster, N., Yates, J.R., Huffaker, T.C., Drubin, D.G., and Barnes, G. (2001). Implication of a novel multiprotein Dam1p complex in outer kinetochore function. J Cell Biol **155,** 1137-1145.

Chen, E.I., Hewel, J., Felding-Habermann, B., and Yates, J.R., 3rd. (2006). Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). Mol Cell Proteomics **5,** 53-56.

Chen, S., and Harmon, A.C. (2006). Advances in plant proteomics. Proteomics **6,** 5504-5516.

Chen, Y., and Periasamy, A. (2007). Localization of protein-protein interactions in live cells using confocal and spectral imaging FRET microscopy. Indian journal of experimental biology **45,** 48-57.

Chern, M., Richter, T., and Ronald, P.C. (2007). Yeast two-hybrid approaches to dissecting the plant defense response. Methods in molecular biology (Clifton, N.J **354,** 79-83.

Churchman, M.L., Brown, M.L., Kato, N., Kirik, V., Hulskamp, M., Inze, D., De Veylder, L., Walker, J.D., Zheng, Z., Oppenheimer, D.G., Gwin, T., Churchman, J., and Larkin, J.C. (2006). SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana. The Plant cell **18,** 3145-3157.

Citovsky, V., Gafni, Y., and Tzfira, T. (2008). Localizing protein-protein interactions by bimolecular fluorescence complementation in planta. Methods **45,** 196-206.

Citovsky, V., Lee, L.Y., Vyas, S., Glick, E., Chen, M.H., Vainstein, A., Gafni, Y., Gelvin, S.B., and Tzfira, T. (2006). Subcellular localization of interacting proteins by bimolecular fluorescence complementation in planta. Journal of molecular biology **362,** 1120-1131.

Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., and Gauthier, J.M. (2004). Functional proteomics mapping of a human signaling pathway. Genome Res **14,** 1324-1332.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics **6,** 439-450.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., and Sanchez, J.C. (2000). The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis **21,** 1104-1115.

Cox, J., and Mann, M. (2007). Is proteomics the new genomics? Cell **130,** 395-398.

Criqui, M.C., Weingartner, M., Capron, A., Parmentier, Y., Shen, W.H., Heberle-Bors, E., Bogre, L., and Genschik, P. (2001). Sub-cellular localisation of GFP-tagged tobacco mitotic cyclins during the cell cycle and after spindle checkpoint activation. Plant J **28,** 569-581.

Cui, J., Li, P., Li, G., Xu, F., Zhao, C., Li, Y., Yang, Z., Wang, G., Yu, Q., and Shi, T. (2008). AtPID: Arabidopsis thaliana protein interactome database--an integrative platform for plant systems biology. Nucleic acids research **36,** D999-1008.

de Folter, S., Immink, R.G., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. The Plant cell **17,** 1424-1433.

De Schutter, K., Joubes, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S., Inze, D., and De Veylder, L. (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. The Plant cell **19,** 211-225.

De Smet, I., Vassileva, V., De Rybel, B., Levesque, M.P., Grunewald, W., Van Damme, D., Van Noorden, G., Naudts, M., Van Isterdael, G., De Clercq, R., Wang, J.Y., Meuli, N., Vanneste, S., Friml, J., Hilson, P., Jürgens, G., Ingram, G.I., Inzé, D., Benfey, P.N., and Beeckman, T. (2008). Receptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root. Science **in press**.

De Veylder, L., Beeckman, T., and Inze, D. (2007). The ins and outs of the plant cell cycle. Nature reviews **8,** 655-665.

De Veylder, L., Segers, G., Glab, N., Van Montagu, M., and Inze, D. (1997a). Identification of proteins interacting with the Arabidopsis Cdc2aAt protein. Journal of experimental botany **48,** 2113-2114.

De Veylder, L., Segers, G., Glab, N., Casteels, P., Van Montagu, M., and Inze, D. (1997b). The Arabidopsis Cks1At protein binds the cyclin-dependent kinases Cdc2aAt and Cdc2bAt. FEBS Lett **412,** 446-452.

De Veylder, L., Beeckman, T., Beemster, G.T., Krols, L., Terras, F., Landrieu, I., van der Schueren, E., Maes, S., Naudts, M., and Inze, D. (2001). Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. The Plant cell **13,** 1653-1668.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A **105,** 10762-10767.

Dewitte, W., and Murray, J.A. (2003). The plant cell cycle. Annu Rev Plant Biol **54,** 235-264.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C., Nieuwland, J., Prinsen, E., Sundaresan, V., and Murray, J.A. (2007). Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. Proc Natl Acad Sci U S A **104,** 14537-14542.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell **86,** 423-433.

Drakas, R., Prisco, M., and Baserga, R. (2005). A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. Proteomics **5,** 132-137.

Drewes, G., and Bouwmeester, T. (2003). Global approaches to protein-protein interactions. Current opinion in cell biology **15,** 199-205.

Dudkina, N.V., Eubel, H., Keegstra, W., Boekema, E.J., and Braun, H.P. (2005). Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. Proc Natl Acad Sci U S A **102,** 3225-3229.

Dufresne, P.J., Thivierge, K., Cotton, S., Beauchemin, C., Ide, C., Ubalijoro, E., Laliberte, J.F., and Fortin, M.G. (2008). Heat shock 70 protein interaction with Turnip mosaic virus RNA-dependent RNA polymerase within virus-induced membrane vesicles. Virology **374,** 217-227.

Dunkley, T.P., Dupree, P., Watson, R.B., and Lilley, K.S. (2004). The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in Arabidopsis thaliana. Biochemical Society transactions **32,** 520-523.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J **45,** 616-629.

Ebel, C., Mariconti, L., and Gruissem, W. (2004). Plant retinoblastoma homologues control nuclear proliferation in the female gametophyte. Nature **429,** 776-780.

Eeckhout, D., De Clercq, A., Van De Slijke, E., Van Leene, J., Stals, H., Casteels, P., Persiau, G., Vercammen, D., Van Breusegem, F., Zabeau, M., Inze, D., Jespers, L., Depicker, A., and De Jaeger, G. (2004). A technology platform for the fast production of monoclonal recombinant antibodies against plant proteins and peptides. J Immunol Methods **294,** 181-187.

Ehrhard, K.N., Jacoby, J.J., Fu, X.Y., Jahn, R., and Dohlman, H.G. (2000). Use of G-protein fusions to monitor integral membrane protein-protein interactions in yeast. Nature biotechnology **18,** 1075-1079.

Ekins, R.P. (1989). Multi-analyte immunoassay. Journal of pharmaceutical and biomedical analysis **7,** 155-168.

Ellmark, P., Ohlin, M., Borrebaeck, C.A., and Furebring, C. (2004). A novel mammalian display system for the selection of protein-protein interactions by decoy receptor engagement. J Mol Recognit **17,** 316-322.

Erhardt, M., Stoppin-Mellet, V., Campagne, S., Canaday, J., Mutterer, J., Fabian, T., Sauter, M., Muller, T., Peter, C., Lambert, A.M., and Schmit, A.C. (2002). The plant Spc98p homologue colocalizes with gamma-tubulin at microtubule nucleation sites and is required for microtubule nucleation. J Cell Sci **115,** 2423-2431.

Estevez, A.M., Kempf, T., and Clayton, C. (2001). The exosome of Trypanosoma brucei. The EMBO journal **20,** 3831-3839.

Eubel, H., Braun, H.P., and Millar, A.H. (2005). Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. Plant Methods **1,** 11.

Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y.V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J.P., Duewel, H.S., Stewart, II, Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S.L., Moran, M.F., Morin, G.B., Topaloglou, T., and Figeys, D. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol **3,** 89.

Fang, Y., Macool, D.J., Xue, Z., Heppard, E.P., Hainey, C.F., Tingey, S.V., and Miao, G.H. (2002). Development of a high-throughput yeast two-hybrid screening system to study protein-protein interactions in plants. Mol Genet Genomics **267,** 142-153.

Feilner, T., and Kersten, B. (2007). Phosphorylation studies using plant protein microarrays. Methods in molecular biology (Clifton, N.J **355,** 379-390.

Feilner, T., Hultschig, C., Lee, J., Meyer, S., Immink, R.G., Koenig, A., Possling, A., Seitz, H., Beveridge, A., Scheel, D., Cahill, D.J., Lehrach, H., Kreutzberger, J., and Kersten, B. (2005). High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. Mol Cell Proteomics **4,** 1558-1568.

Feng, X., Liu, X., Luo, Q., and Liu, B.F. (2008). Mass spectrometry in systems biology: An overview. Mass spectrometry reviews **27,** 635-660.

Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nature biotechnology **20,** 301-305.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature **340,** 245-246.

Figeys, D. (2002). Functional proteomics: mapping protein-protein interactions and pathways. Current opinion in molecular therapeutics **4,** 210-215.

Fleury, D., Himanen, K., Cnops, G., Nelissen, H., Boccardi, T.M., Maere, S., Beemster, G.T., Neyt, P., Anami, S., Robles, P., Micol, J.L., Inze, D., and Van Lijsebettens, M. (2007). The Arabidopsis thaliana homolog of yeast BRE1 has a function in cell cycle regulation during early leaf and root growth. The Plant cell **19,** 417-432.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nature biotechnology **21,** 89-92.

Forment, J., Naranjo, M.A., Roldan, M., Serrano, R., and Vicente, O. (2002). Expression of Arabidopsis SR-like splicing proteins confers salt tolerance to yeast and transgenic plants. Plant J **30,** 511-519.

Forreiter, C., Kirschner, M., and Nover, L. (1997). Stable transformation of an Arabidopsis cell suspension culture with firefly luciferase providing a cellular system for analysis of chaperone activity in vivo. The Plant cell **9,** 2171-2181.

Francis, D. (2007). The plant cell cycle--15 years on. New Phytol **174,** 261-278.

Fromont-Racine, M., Rain, J.C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature genetics **16,** 277-282.

Fujikawa, Y., and Kato, N. (2007). Split luciferase complementation assay to study protein-protein interactions in Arabidopsis protoplasts. Plant J **52,** 185-195.

Fulop, K., Pettko-Szandtner, A., Magyar, Z., Miskolczi, P., Kondorosi, E., Dudits, D., and Bako, L. (2005). The Medicago CDKC;1-CYCLINT;1 kinase complex phosphorylates the carboxy-terminal domain of RNA polymerase II and promotes transcription. Plant J **42,** 810-820.

Garavelli, J.S. (2004). The RESID Database of Protein Modifications as a resource and annotation tool. Proteomics **4,** 1527-1533.

Gasch, A.P., and Eisen, M.B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. Genome biology **3,** RESEARCH0059.

Gavin, A.C., and Superti-Furga, G. (2003). Protein complexes and proteome organization from yeast to man. Current opinion in chemical biology **7,** 21-27.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., and Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature **440,** 631-636.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino,

A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415,** 141-147.

Ge, H., Walhout, A.J., and Vidal, M. (2003). Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet **19,** 551-560.

Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M. (2007). A predicted interactome for Arabidopsis. Plant Physiol **145,** 317-329.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences of the United States of America **100,** 6940-6945.

Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nature biotechnology **21,** 566-569.

Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nature reviews **8,** 645-654.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C.A., Finley, R.L., Jr., White, K.P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R.A., McKenna, M.P., Chant, J., and Rothberg, J.M. (2003). A protein interaction map of Drosophila melanogaster. Science **302,** 1727-1736.

Gloeckner, C.J., Boldt, K., Schumacher, A., Roepman, R., and Ueffing, M. (2007). A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. Proteomics **7,** 4228-4234.

Gong, W., Shen, Y.P., Ma, L.G., Pan, Y., Du, Y.L., Wang, D.H., Yang, J.Y., Hu, L.D., Liu, X.F., Dong, C.X., Ma, L., Chen, Y.H., Yang, X.Y., Gao, Y., Zhu, D., Tan, X., Mu, J.Y., Zhang, D.B., Liu, Y.L., Dinesh-Kumar, S.P., Li, Y., Wang, X.P., Gu, H.Y., Qu, L.J., Bai, S.N., Lu, Y.T., Li, J.Y., Zhao, J.D., Zuo, J., Huang, H., Deng, X.W., and Zhu, Y.X. (2004). Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant physiology **135,** 773-782.

Gould, K.L., Ren, L., Feoktistova, A.S., Jennings, J.L., and Link, A.J. (2004). Tandem affinity purification and identification of protein complex components. Methods **33,** 239-244.

Graumann, J., Dunipace, L.A., Seol, J.H., McDonald, W.H., Yates, J.R., 3rd, Wold, B.J., and Deshaies, R.J. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. Mol Cell Proteomics **3,** 226-237.

Gruhler, A., Schulze, W.X., Matthiesen, R., Mann, M., and Jensen, O.N. (2005). Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry. Mol Cell Proteomics **4,** 1697-1709.

Guerrero, C., Tagwerker, C., Kaiser, P., and Huang, L. (2006). An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. Mol Cell Proteomics **5,** 366-378.

Gunsalus, K.C., Ge, H., Schetter, A.J., Goldberg, D.S., Han, J.D., Hao, T., Berriz, G.F., Bertin, N., Huang, J., Chuang, L.S., Li, N., Mani, R., Hyman, A.A., Sonnichsen, B., Echeverri, C.J., Roth, F.P., Vidal, M., and Piano, F. (2005). Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. Nature **436,** 861-865.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proceedings of the National Academy of Sciences of the United States of America **97,** 9390-9395.

Hack, C.J. (2004). Integrated transcriptome and proteome data: the challenges ahead. Briefings in functional genomics & proteomics **3,** 212-219.

Hartman, N.T., Sicilia, F., Lilley, K.S., and Dupree, P. (2007). Proteomic complex detection using sedimentation. Analytical chemistry **79,** 2078-2083.

Hartung, F., Suer, S., Bergmann, T., and Puchta, H. (2006). The role of AtMUS81 in DNA repair and its genetic interaction with the helicase AtRecQ4A. Nucleic acids research **34,** 4438-4448.

Hase, Y., Trung, K.H., Matsunaga, T., and Tanaka, A. (2006). A mutation in the uvi4 gene promotes progression of endo-reduplication and confers increased tolerance towards ultraviolet B light. Plant J **46,** 317-326.

Hayes, M.J., Kimata, Y., Wattam, S.L., Lindon, C., Mao, G., Yamano, H., and Fry, A.M. (2006). Early mitotic degradation of Nek2A depends on Cdc20-independent interaction with the APC/C. Nat Cell Biol **8,** 607-614.

Healy, J.M., Menges, M., Doonan, J.H., and Murray, J.A. (2001). The Arabidopsis D-type cyclins CycD2 and CycD3 both interact in vivo with the PSTAIRE cyclin-dependent kinase Cdc2a but are differentially controlled. J Biol Chem **276,** 7041-7047.

Hearn, M.T., and Acosta, D. (2001). Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins. J Mol Recognit **14,** 323-369.

Heazlewood, J.L., Tonti-Filippini, J., Verboom, R.E., and Millar, A.H. (2005). Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. Plant physiology **139,** 598-609.

Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., Small, I., and Millar, A.H. (2007). SUBA: the Arabidopsis Subcellular Database. Nucleic acids research **35,** D213-218.

Hemerly, A., Engler Jde, A., Bergounioux, C., Van Montagu, M., Engler, G., Inze, D., and Ferreira, P. (1995). Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. The EMBO journal **14,** 3925-3936.

Hilson, P. (2006). Cloned sequence repertoires for small- and large-scale biology. Trends in plant science **11,** 133-141.

Hink, M.A., Bisselin, T., and Visser, A.J. (2002). Imaging protein-protein interactions in living cells. Plant molecular biology **50,** 871-883.

Hirano, H., Islam, N., and Kawasaki, H. (2004). Technical aspects of functional proteomics in plants. Phytochemistry **65,** 1487-1498.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D., and Tyers, M. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature **415,** 180-183.

Hochholdinger, F., Sauer, M., Dembinsky, D., Hoecker, N., Muthreich, N., Saleem, M., and Liu, Y. (2006). Proteomic dissection of plant development. Proteomics **6,** 4076-4083.

Honey, S., Schneider, B.L., Schieltz, D.M., Yates, J.R., and Futcher, B. (2001). A novel multiple affinity purification tag and its use in identification of proteins associated with a cyclin-CDK complex. Nucleic acids research **29,** E24.

Hu, C.D., Chinenov, Y., and Kerppola, T.K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Molecular cell **9,** 789-798.

Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., and Rhee, S.Y. (2001). The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic acids research **29,** 102-105.

Huanca-Mamani, W., Garcia-Aguilar, M., Leon-Martinez, G., Grossniklaus, U., and Vielle-Calzada, J.P. (2005). CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in Arabidopsis thaliana. Proc Natl Acad Sci U S A **102,** 17231-17236.

Imai, K.K., Ohashi, Y., Tsuge, T., Yoshizumi, T., Matsui, M., Oka, A., and Aoyama, T. (2006). The A-type cyclin CYCA2;3 is a key regulator of ploidy levels in Arabidopsis endoreduplication. The Plant cell **18,** 382-396.

Inze, D. (2005). Green light for the cell cycle. The EMBO journal **24,** 657-662.

Inze, D., and Veylder, L.D. (2006). Cell Cycle Regulation in Plant Development. Annu Rev Genet.

Inze, D., and De Veylder, L. (2006). Cell cycle regulation in plant development. Annu Rev Genet **40,** 77-105.

Isaacson, T., Damasceno, C.M., Saravanan, R.S., He, Y., Catala, C., Saladie, M., and Rose, J.K. (2006). Sample extraction techniques for enhanced proteomic analysis of plant tissues. Nat Protoc **1,** 769-774.

Ishida, S., Huang, E., Zuzan, H., Spang, R., Leone, G., West, M., and Nevins, J.R. (2001). Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. Mol Cell Biol **21,** 4684-4699.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A **98,** 4569-4574.

Jacqmard, A., De Veylder, L., Segers, G., de Almeida Engler, J., Bernier, G., Van Montagu, M., and Inze, D. (1999). Expression of CKS1At in Arabidopsis thaliana indicates a role for the protein in both the mitotic and the endoreduplication cycle. Planta **207,** 496-504.

Jensen, L.J., Jensen, T.S., de Lichtenberg, U., Brunak, S., and Bork, P. (2006). Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature **443,** 594-597.

Jorrin, J.V., Maldonado, A.M., and Castillejo, M.A. (2007). Plant proteome analysis: A 2006 update. Proteomics **7,** 2947-2962.

Junttila, M.R., Saarinen, S., Schmidt, T., Kast, J., and Westermarck, J. (2005). Single-step Strep-tag purification for the isolation and identification of protein complexes from mammalian cells. Proteomics **5,** 1199-1203.

Karimi, M., De Meyer, B., and Hilson, P. (2005). Modular cloning in plant cells. Trends in plant science **10,** 103-105.

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. Plant Physiol **145,** 1144-1154.

Karlova, R., Boeren, S., Russinova, E., Aker, J., Vervoort, J., and de Vries, S. (2006). The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. The Plant cell **18,** 626-638.

Karp, N.A., Kreil, D.P., and Lilley, K.S. (2004). Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. Proteomics **4,** 1421-1432.

Kawamura, K., Murray, J.A., Shinmyo, A., and Sekine, M. (2006). Cell cycle regulated D3-type cyclins form active complexes with plant-specific B-type cyclin-dependent kinase in vitro. Plant molecular biology **61,** 311-327.

Keefe, A.D., Wilson, D.S., Seelig, B., and Szostak, J.W. (2001). One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. Protein Expr Purif **23,** 440-446.

Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., and Hermjakob, H. (2007). IntAct--open source resource for molecular interaction data. Nucleic acids research **35,** D561-565.

Kersten, B., and Feilner, T. (2007). Generation of plant protein microarrays and investigation of antigen-antibody interactions. Methods in molecular biology (Clifton, N.J **355,** 365-378.

Kersten, B., Agrawal, G.K., Iwahashi, H., and Rakwal, R. (2006). Plant phosphoproteomics: a long road ahead. Proteomics **6,** 5517-5528.

Kersten, B., Burkle, L., Kuhn, E.J., Giavalisco, P., Konthur, Z., Lueking, A., Walter, G., Eickhoff, H., and Schneider, U. (2002). Large-scale plant proteomics. Plant molecular biology **48,** 133-141.

Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D., and Nam, H.G. (2008). Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. Nature **455,** 1134-1137.

Kim, W.Y., Fujiwara, S., Suh, S.S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H.G., and Somers, D.E. (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. Nature **449,** 356-360.

Kimbara, J., Endo, T.R., and Nasuda, S. (2004). Characterization of the genes encoding for MAD2 homologues in wheat. Chromosome Res **12,** 703-714.

Kimura, A., Adachi, N., and Horikoshi, M. (2003). Series of vectors to evaluate the position and the order of two different affinity tags for purification of protein complexes. Anal Biochem **314,** 253-259.

Kitsios, G., Alexiou, K.G., Bush, M., Shaw, P., and Doonan, J.H. (2008). A cyclin-dependent protein kinase, CDKC2, colocalizes with and modulates the distribution of spliceosomal components in Arabidopsis. Plant J **54,** 220-235.

Kjell, J., Rasmusson, A.G., Larsson, H., and Widell, S. (2004). Protein complexes of the plant plasma membrane resolved by Blue Native PAGE. Physiologia Plantarum **121,** 546-555.

Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X. (2003). Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. Mol Cell Proteomics **2,** 1225-1233.

Kono, A., Ohno, R., Umeda-Hara, C., Uchimiya, H., and Umeda, M. (2006). A distinct type of cyclin D, CYCD4;2, involved in the activation of cell division in Arabidopsis. Plant Cell Rep **25,** 540-545.

Kono, A., Umeda-Hara, C., Lee, J., Ito, M., Uchimiya, H., and Umeda, M. (2003). Arabidopsis D-type cyclin CYCD4;1 is a novel cyclin partner of B2-type cyclin-dependent kinase. Plant Physiol **132,** 1315-1321.

Koroleva, O.A., Tomlinson, M.L., Leader, D., Shaw, P., and Doonan, J.H. (2005). High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. Plant J **41,** 162-174.

Krempler, A., Kartarius, S., Gunther, J., and Montenarh, M. (2005). Cyclin H is targeted to the nucleus by C-terminal nuclear localization sequences. Cell Mol Life Sci **62,** 1379-1387.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., and Greenblatt, J.F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature **440,** 637-643.

Kung, L.A., and Snyder, M. (2006). Proteome chips for whole-organism assays. Nature reviews **7,** 617-622.

Kurucz, E., Ando, I., Sumegi, M., Holzl, H., Kapelari, B., Baumeister, W., and Udvardy, A. (2002). Assembly of the Drosophila 26 S proteasome is accompanied by extensive subunit rearrangements. The Biochemical journal **365,** 527-536.

Kwee, H.S., and Sundaresan, V. (2003). The NOMEGA gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in Arabidopsis. Plant J **36,** 853-866.

Kwon, C.S., Hibara, K., Pfluger, J., Bezhani, S., Metha, H., Aida, M., Tasaka, M., and Wagner, D. (2006). A role for chromatin remodeling in regulation of CUC gene expression in the Arabidopsis cotyledon boundary. Development **133,** 3223-3230.

Kyriakakis, P., Tipping, M., Abed, L., and Veraksa, A. (2008). Tandem affinity purification in Drosophila: The advantages of the GS-TAP system. Fly (Austin) **2**.

Lammens, T., Boudolf, V., Kheibarshekan, L., Zalmas, L.P., Gaamouche, T., Maes, S., Vanstraelen, M., Kondorosi, E., La Thangue, N.B., Govaerts, W., Inze, D., and De Veylder, L. (2008). Atypical E2F activity restrains APC/CCCS52A2 function obligatory for endocycle onset. Proc Natl Acad Sci U S A **105,** 14721-14726.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jorgensen, T.J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics **4,** 873-886.

Lee, K.M., Miklos, I., Du, H., Watt, S., Szilagyi, Z., Saiz, J.E., Madabhushi, R., Penkett, C.J., Sipiczki, M., Bahler, J., and Fisher, R.P. (2005). Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression in fission yeast. Mol Biol Cell **16,** 2734-2745.

Li, J., Li, X., Su, H., Chen, H., and Galbraith, D.W. (2006a). A framework of integrating gene relations from heterogeneous data sources: an experiment on Arabidopsis thaliana. Bioinformatics (Oxford, England) **22,** 2037-2043.

Li, S., Ehrhardt, D.W., and Rhee, S.Y. (2006b). Systematic analysis of Arabidopsis organelles and a protein localization database for facilitating fluorescent tagging of full-length Arabidopsis proteins. Plant physiology **141,** 527-539.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E., and Vidal, M. (2004). A map of the interactome network of the metazoan C. elegans. Science **303,** 540-543.

Littlepage, L.E., and Ruderman, J.V. (2002). Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev **16,** 2274-2285.

Liu, J., Zhang, Y., Qin, G., Tsuge, T., Sakaguchi, N., Luo, G., Sun, K., Shi, D., Aki, S., Zheng, N., Aoyama, T., Oka, A., Yang, W., Umeda, M., Xie, Q., Gu, H., and Qu, L.J. (2008). Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during Arabidopsis gametogenesis. The Plant cell **20,** 1538-1554.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature biotechnology **25,** 117-124.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics (Oxford, England) **21,** 3448-3449.

Magyar, Z., Atanassova, A., De Veylder, L., Rombauts, S., and Inze, D. (2000). Characterization of two distinct DP-related genes from Arabidopsis thaliana. FEBS Lett **486,** 79-87.

Magyar, Z., De Veylder, L., Atanassova, A., Bako, L., Inze, D., and Bogre, L. (2005). The role of the Arabidopsis E2FB transcription factor in regulating auxin-dependent cell division. The Plant cell **17,** 2527-2541.

Magyar, Z., Meszaros, T., Miskolczi, P., Deak, M., Feher, A., Brown, S., Kondorosi, E., Athanasiadis, A., Pongor, S., Bilgin, M., Bako, L., Koncz, C., and Dudits, D. (1997). Cell cycle phase specificity of putative cyclin-dependent kinase variants in synchronized alfalfa cells. The Plant cell **9,** 223-235.

Maor, R., Jones, A., Nuhse, T.S., Studholme, D.J., Peck, S.C., and Shirasu, K. (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics **6,** 601-610.

Mayer, M.L., Gygi, S.P., Aebersold, R., and Hieter, P. (2001). Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. Molecular cell **7,** 959-970.

Menges, M., and Murray, J.A. (2002). Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J **30,** 203-212.

Menges, M., and Murray, J.A. (2004). Cryopreservation of transformed and wild-type Arabidopsis and tobacco cell suspension cultures. Plant J **37,** 635-644.

Menges, M., and Murray, J.A. (2006). Synchronization, transformation, and cryopreservation of suspension-cultured cells. Methods in molecular biology (Clifton, N.J **323,** 45-61.

Menges, M., Hennig, L., Gruissem, W., and Murray, J.A. (2003). Genome-wide gene expression in an Arabidopsis cell suspension. Plant molecular biology **53,** 423-442.

Menges, M., de Jager, S.M., Gruissem, W., and Murray, J.A. (2005). Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. Plant J **41,** 546-566.

Menon, S., Rubio, V., Wang, X., Deng, X.W., and Wei, N. (2005). Purification of the COP9 signalosome from porcine spleen, human cell lines, and Arabidopsis thaliana plants. Methods in enzymology **398,** 468-481.

Michnick, S.W., Remy, I., Campbell-Valois, F.X., Vallee-Belisle, A., and Pelletier, J.N. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. Methods in enzymology **328,** 208-230.

Miernyk, J.A., and Thelen, J.J. (2008). Biochemical approaches for discovering protein-protein interactions. Plant J **53,** 597-609.

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. Genomics **85,** 1-15.

Morsy, M., Gouthu, S., Orchard, S., Thorneycroft, D., Harper, J.F., Mittler, R., and Cushman, J.C. (2008). Charting plant interactomes: possibilities and challenges. Trends in plant science **13,** 183-191.

Muller, S., Smertenko, A., Wagner, V., Heinrich, M., Hussey, P.J., and Hauser, M.T. (2004). The plant microtubule-associated protein AtMAP65-3/PLE is essential for cytokinetic phragmoplast function. Curr Biol **14,** 412-417.

Nakagami, H., Sekine, M., Murakami, H., and Shinmyo, A. (1999). Tobacco retinoblastoma-related protein phosphorylated by a distinct cyclin-dependent kinase complex with Cdc2/cyclin D in vitro. Plant J **18,** 243-252.

Nakai, T., Kato, K., Shinmyo, A., and Sekine, M. (2006). Arabidopsis KRPs have distinct inhibitory activity toward cyclin D2-associated kinases, including plant-specific B-type cyclin-dependent kinase. FEBS Lett **580,** 336-340.

Nelson, B.K., Cai, X., and Nebenfuhr, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J **51,** 1126-1136.

Neufeld, T.P., de la Cruz, A.F., Johnston, L.A., and Edgar, B.A. (1998). Coordination of growth and cell division in the Drosophila wing. Cell **93,** 1183-1193.

Nolan, T., Hands, R.E., and Bustin, S.A. (2006). Quantification of mRNA using real-time RT-PCR. Nature protocols **1,** 1559-1582.

Nuhse, T.S., Stensballe, A., Jensen, O.N., and Peck, S.C. (2004). Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. The Plant cell **16,** 2394-2405.

Nurse, P. (1994). Ordering S phase and M phase in the cell cycle. Cell **79,** 547-550.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Shibata, D., Saito, K., and Ohta, H. (2007). ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. Nucleic acids research **35,** D863-869.

Oda, Y., Nagasu, T., and Chait, B.T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nature biotechnology **19,** 379-382.

Ohad, N., Shichrur, K., and Yalovsky, S. (2007). The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. Plant Physiol **145,** 1090-1099.

Ong, S.E., and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nature protocols **1,** 2650-2660.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics **1,** 376-386.

Ooi, S.L., Shoemaker, D.D., and Boeke, J.D. (2003). DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nature genetics **35,** 277-286.

Ooi, S.L., Pan, X., Peyser, B.D., Ye, P., Meluh, P.B., Yuan, D.S., Irizarry, R.A., Bader, J.S., Spencer, F.A., and Boeke, J.D. (2006). Global synthetic-lethality analysis and yeast functional profiling. Trends Genet **22,** 56-63.

Pan, X., Yuan, D.S., Ooi, S.L., Wang, X., Sookhai-Mahadeo, S., Meluh, P., and Boeke, J.D. (2007). dSLAM analysis of genome-wide genetic interactions in Saccharomyces cerevisiae. Methods (San Diego, Calif **41,** 206-221.

Park, O.K. (2004). Proteomic studies in plants. Journal of biochemistry and molecular biology **37,** 133-138.

Paz-Ares, J., and The Regia, C. (2002). REGIA, An EU Project on Functional Genomics of Transcription Factors From Arabidopsis Thaliana. Comp Funct Genomics **3,** 102-108.

Peck, S.C. (2005). Update on proteomics in Arabidopsis. Where do we go from here? Plant physiology **138,** 591-599.

Peck, S.C. (2006). Analysis of protein phosphorylation: methods and strategies for studying kinases and substrates. Plant J **45,** 512-522.

Pelletier, J.N., Campbell-Valois, F.X., and Michnick, S.W. (1998). Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proceedings of the National Academy of Sciences of the United States of America **95,** 12141-12146.

Pepper, L.R., Cho, Y.K., Boder, E.T., and Shusta, E.V. (2008). A decade of yeast surface display technology: where are we now? Comb Chem High Throughput Screen **11,** 127-134.

Peres, A., Churchman, M.L., Hariharan, S., Himanen, K., Verkest, A., Vandepoele, K., Magyar, Z., Hatzfeld, Y., Van Der Schueren, E., Beemster, G.T., Frankard, V., Larkin, J.C., Inze, D., and De Veylder, L. (2007). Novel plant-specific cyclin-dependent kinase inhibitors induced by biotic and abiotic stresses. J Biol Chem **282,** 25588-25596.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20,** 3551-3567.

Peters, J.M. (1998). SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. Current opinion in cell biology **10,** 759-768.

Peters, J.M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. Nature reviews **7,** 644-656.

Pinkse, M.W., Uitto, P.M., Hilhorst, M.J., Ooms, B., and Heck, A.J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Analytical chemistry **76,** 3935-3943.

Planchais, S., Samland, A.K., and Murray, J.A. (2004). Differential stability of Arabidopsis D-type cyclins: CYCD3;1 is a highly unstable protein degraded by a proteasome-dependent mechanism. Plant J **38,** 616-625.

Popescu, S.C., Popescu, G.V., Bachan, S., Zhang, Z., Seay, M., Gerstein, M., Snyder, M., and Dinesh-Kumar, S.P. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. Proc Natl Acad Sci U S A **104,** 4730-4735.

Porceddu, A., Stals, H., Reichheld, J.P., Segers, G., De Veylder, L., Barroco, R.P., Casteels, P., Van Montagu, M., Inze, D., and Mironov, V. (2001). A plant-specific cyclin-dependent kinase is involved in the control of G2/M progression in plants. J Biol Chem **276,** 36354-36360.

Portereiko, M.F., Sandaklie-Nikolova, L., Lloyd, A., Dever, C.A., Otsuga, D., and Drews, G.N. (2006). NUCLEAR FUSION DEFECTIVE1 encodes the Arabidopsis RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. Plant Physiol **141,** 957-965.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods **24,** 218-229.

Ramirez-Parra, E., Frundt, C., and Gutierrez, C. (2003). A genome-wide identification of E2F-regulated genes in Arabidopsis. Plant J **33,** 801-811.

Ramirez-Parra, E., Lopez-Matas, M.A., Frundt, C., and Gutierrez, C. (2004). Role of an atypical E2F transcription factor in the control of Arabidopsis cell growth and differentiation. The Plant cell **16,** 2350-2363.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T., Hudson, J.R., Jr., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S., Endress, G.A., Jenna, S., Chevet, E., Papasotiropoulos, V., Tolias, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H., Doucette-Stamm, L., Hill, D.E., and Vidal, M. (2003). C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nature genetics **34,** 35-41.

Rechsteiner, M., and Rogers, S.W. (1996). PEST sequences and regulation by proteolysis. Trends Biochem Sci **21,** 267-271.

Remy, I., Campbell-Valois, F.X., and Michnick, S.W. (2007a). Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. Nat Protoc **2,** 2120-2125.

Remy, I., Ghaddar, G., and Michnick, S.W. (2007b). Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions. Nat Protoc **2,** 2302-2306.

Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D.C., Wu, Y., Xu, I., Yoo, D., Yoon, J., and Zhang, P. (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated

gateway to Arabidopsis biology, research materials and community. Nucleic acids research **31,** 224-228.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16,** 276-277.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology **17,** 1030-1032.

Riou-Khamlichi, C., Huntley, R., Jacqmard, A., and Murray, J.A. (1999). Cytokinin activation of Arabidopsis cell division through a D-type cyclin. Science **283,** 1541-1544.

Rodi, D.J., and Makowski, L. (1999). Phage-display technology--finding a needle in a vast molecular haystack. Current opinion in biotechnology **10,** 87-93.

Rohila, J.S., Chen, M., Cerny, R., and Fromm, M.E. (2004). Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J **38,** 172-181.

Rohila, J.S., Chen, M., Chen, S., Chen, J., Cerny, R., Dardick, C., Canlas, P., Xu, X., Gribskov, M., Kanrar, S., Zhu, J.K., Ronald, P., and Fromm, M.E. (2006). Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. Plant J **46,** 1-13.

Rose, J.K.C., Bashir, S., Giovannoni, J.J., Jahn, M.M., and Saravanan, R.S. (2004). Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant J **39,** 715-733.

Rossignol, M., Peltier, J.B., Mock, H.P., Matros, A., Maldonado, A.M., and Jorrin, J.V. (2006). Plant proteome analysis: a 2004-2006 update. Proteomics **6,** 5529-5548.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., and Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature **437,** 1173-1178.

Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S.P., and Deng, X.W. (2005). An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. Plant J **41,** 767-778.

Schaff, J.E., Mbeunkui, F., Blackburn, K., Bird, D.M., and Goshe, M.B. (2008). SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. Plant J.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (New York, N.Y **270,** 467-470.

Schmidt, U., Wollmann, Y., Franke, C., Grosse, F., Saluz, H.P., and Hanel, F. (2008). Characterization of the interaction between the human DNA topoisomerase IIbeta-binding protein 1 (TopBP1) and the cell division cycle 45 (Cdc45) protein. The Biochemical journal **409,** 169-177.

Scott, J.K., and Smith, G.P. (1990). Searching for peptide ligands with an epitope library. Science **249,** 386-390.

Seifert, G.J. (2004). Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Curr Opin Plant Biol **7,** 277-284.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13,** 2498-2504.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Analytical chemistry **68,** 850-858.

Shevchenko, A., Schaft, D., Roguev, A., Pijnappel, W.W., Stewart, A.F., and Shevchenko, A. (2002). Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective. Mol Cell Proteomics **1,** 204-212.

Shimotohno, A., Umeda-Hara, C., Bisova, K., Uchimiya, H., and Umeda, M. (2004). The plant-specific kinase CDKF;1 is involved in activating phosphorylation of cyclin-dependent kinase-activating kinases in Arabidopsis. The Plant cell **16,** 2954-2966.

Shimotohno, A., Ohno, R., Bisova, K., Sakaguchi, N., Huang, J., Koncz, C., Uchimiya, H., and Umeda, M. (2006). Diverse phosphoregulatory mechanisms controlling cyclin-dependent kinase-activating kinases in Arabidopsis. Plant J **47,** 701-710.

Shoemaker, B.A., and Panchenko, A.R. (2007). Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol **3,** e43.

Shultz, R.W., Tatineni, V.M., Hanley-Bowdoin, L., and Thompson, W.F. (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. Plant Physiol **144,** 1697-1714.

Shyu, Y.J., Suarez, C.D., and Hu, C.D. (2008). Visualization of ternary complexes in living cells by using a BiFC-based FRET assay. Nat Protoc **3,** 1693-1702.

Skerra, A., and Schmidt, T.G. (2000). Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins. Methods in enzymology **326,** 271-304.

Smith, D.B., and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67,** 31-40.

Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proceedings of the National Academy of Sciences of the United States of America **95,** 5187-5192.

Stals, H., Bauwens, S., Traas, J., Van Montagu, M., Engler, G., and Inze, D. (1997). Plant CDC2 is not only targeted to the pre-prophase band, but also co-localizes with the spindle, phragmoplast, and chromosomes. FEBS Lett **418,** 229-234.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E.E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell **122,** 957-968.

Stingl, K., Schauer, K., Ecobichon, C., Labigne, A., Lenormand, P., Rousselle, J.C., Namane, A., and de Reuse, H. (2008). In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification. Mol Cell Proteomics.

Subramaniam, R., Desveaux, D., Spickler, C., Michnick, S.W., and Brisson, N. (2001). Direct visualization of protein interactions in plant cells. Nature biotechnology **19,** 769-772.

Subramanian, C., Xu, Y., Johnson, C.H., and von Arnim, A.G. (2004). In vivo detection of protein-protein interaction in plant cells using BRET. Methods in molecular biology (Clifton, N.J **284,** 271-286.

Swaminathan, K., Yang, Y., Grotz, N., Campisi, L., and Jack, T. (2000). An enhancer trap line associated with a D-class cyclin gene in Arabidopsis. Plant Physiol **124,** 1658-1667.

Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L., and Kaiser, P. (2006). A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol Cell Proteomics **5,** 737-748.

Takahashi, N., Lammens, T., Boudolf, V., Maes, S., Yoshizumi, T., De Jaeger, G., Witters, E., Inze, D., and De Veylder, L. (2008). The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. The EMBO journal **27,** 1840-1851.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science **320,** 1465-1470.

Tardif, G., Kane, N.A., Adam, H., Labrie, L., Major, G., Gulick, P., Sarhan, F., and Laliberte, J.F. (2007). Interaction network of proteins associated with abiotic stress response and development in wheat. Plant molecular biology **63,** 703-718.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology **60,** 523-533.

Thaminy, S., Miller, J., and Stagljar, I. (2004). The split-ubiquitin membrane-based yeast two-hybrid system. Methods in molecular biology (Clifton, N.J **261,** 297-312.

Tian, G.W., Mohanty, A., Chary, S.N., Li, S., Paap, B., Drakakaki, G., Kopec, C.D., Li, J., Ehrhardt, D., Jackson, D., Rhee, S.Y., Raikhel, N.V., and Citovsky, V. (2004). High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. Plant physiology **135,** 25-38.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics **1,** 377-396.

Torres Acosta, J.A., de Almeida Engler, J., Raes, J., Magyar, Z., De Groodt, R., Inze, D., and De Veylder, L. (2004). Molecular characterization of Arabidopsis PHO80-like proteins, a novel class of CDKA;1-interacting cyclins. Cell Mol Life Sci **61,** 1485-1497.

Traas, J., Bellini, C., Nacry, P., Kronenberger, J., Bouchez, D., and Caboche, M. (1995). Normal Differentiation Patterns in Plants Lacking Microtubular Preprophase Bands. Nature **375,** 676-677.

Tsai, A., and Carstens, R.P. (2006). An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. Nat Protoc **1,** 2820-2827.

Tsesmetzis, N., Couchman, M., Higgins, J., Smith, A., Doonan, J.H., Seifert, G.J., Schmidt, E.E., Vastrik, I., Birney, E., Wu, G., D'Eustachio, P., Stein, L.D., Morris, R.J., Bevan, M.W., and Walsh, S.V. (2008). Arabidopsis reactome: a foundation knowledgebase for plant systems biology. The Plant cell **20,** 1426-1436.

Tsugita, A., and Kamo, M. (1999). 2-D electrophoresis of plant proteins. Methods in molecular biology (Clifton, N.J **112,** 95-97.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J.M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403,** 623-627.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M.G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skollermo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L.L., Hober, S., and Ponten, F. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics **4,** 1920-1932.

Umeda, M., Shimotohno, A., and Yamaguchi, M. (2005). Control of cell division and transcription by cyclin-dependent kinase-activating kinases in plants. Plant Cell Physiol **46,** 1437-1442.

Underwood, B.A., Vanderhaeghen, R., Whitford, R., Town, C.D., and Hilson, P. (2006). Simultaneous high-throughput recombinational cloning of open reading frames in closed and open configurations. Plant biotechnology journal **4,** 317-324.

Van Aken, O., Pecenkova, T., van de Cotte, B., De Rycke, R., Eeckhout, D., Fromm, H., De Jaeger, G., Witters, E., Beemster, G.T.S., Inze, D., and Van Breusegem, F. (2007). Mitochondrial type-I

prohibitins of Arabidopsis thaliana are required for supporting proficient meristem development. Plant J **52,** 850-864.

Van Leene, J., Witters, E., Inze, D., and De Jaeger, G. (2008). Boosting tandem affinity purification of plant protein complexes. Trends in plant science **13,** 517-520.

Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., De Slijke, E.V., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N., Henderickx, K., De Vijlder, T., Abdelkrim, A., Pharazyn, A., Van Onckelen, H., Inze, D., Witters, E., and De Jaeger, G. (2007). A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics **6,** 1226-1238.

Vandepoele, K., Casneuf, T., and Van de Peer, Y. (2006). Identification of novel regulatory modules in dicotyledonous plants using expression data and comparative genomics. Genome biology **7,** R103.

Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S., and Inze, D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. The Plant cell **14,** 903-916.

Vandepoele, K., Vlieghe, K., Florquin, K., Hennig, L., Beemster, G.T., Gruissem, W., Van de Peer, Y., Inze, D., and De Veylder, L. (2005). Genome-wide identification of potential plant E2F target genes. Plant Physiol **139,** 316-328.

Vanstraelen, M., Torres Acosta, J.A., De Veylder, L., Inze, D., and Geelen, D. (2004). A plant-specific subclass of C-terminal kinesins contains a conserved a-type cyclin-dependent kinase site implicated in folding and dimerization. Plant Physiol **135,** 1417-1429.

Vasilescu, J., Zweitzig, D.R., Denis, N.J., Smith, J.C., Ethier, M., Haines, D.S., and Figeys, D. (2007). The proteomic reactor facilitates the analysis of affinity-purified proteins by mass spectrometry: application for identifying ubiquitinated proteins in human cells. Journal of proteome research **6,** 298-305.

Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., Wu, G., Birney, E., and Stein, L. (2007). Reactome: a knowledge base of biologic pathways and processes. Genome biology **8,** R39.

Veraksa, A., Bauer, A., and Artavanis-Tsakonas, S. (2005). Analyzing protein complexes in Drosophila with tandem affinity purification-mass spectrometry. Dev Dyn **232,** 827-834.

Verkest, A., Weinl, C., Inze, D., De Veylder, L., and Schnittger, A. (2005a). Switching the cell cycle. Kip-related proteins in plant cell cycle control. Plant Physiol **139,** 1099-1106.

Verkest, A., Manes, C.L., Vercruysse, S., Maes, S., Van Der Schueren, E., Beeckman, T., Genschik, P., Kuiper, M., Inze, D., and De Veylder, L. (2005b). The cyclin-dependent kinase inhibitor KRP2 controls the onset of the endoreduplication cycle during Arabidopsis leaf development through inhibition of mitotic CDKA;1 kinase complexes. The Plant cell **17,** 1723-1736.

Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid systems. Nucleic acids research **27,** 919-929.

Villen, J., and Gygi, S.P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc **3,** 1630-1638.

Vodermaier, H.C., Gieffers, C., Maurer-Stroh, S., Eisenhaber, F., and Peters, J.M. (2003). TPR subunits of the anaphase-promoting complex mediate binding to the activator protein CDH1. Curr Biol **13,** 1459-1468.

von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B., and Bork, P. (2007). STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic acids research **35,** D358-362.

von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., and Bork, P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic acids research **33,** D433-437.

Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. (2000). Protein interaction mapping in C. elegans using proteins involved in vulval development. Science **287,** 116-122.

Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K., and Kudla, J. (2004). Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J **40,** 428-438.

Wang, G., Kong, H., Sun, Y., Zhang, X., Zhang, W., Altman, N., DePamphilis, C.W., and Ma, H. (2004). Genome-wide analysis of the cyclin family in Arabidopsis and comparative phylogenetic analysis of plant cyclin-like proteins. Plant Physiol **135,** 1084-1099.

Wang, W., and Chen, X. (2004). HUA ENHANCER3 reveals a role for a cyclin-dependent protein kinase in the specification of floral organ identity in Arabidopsis. Development **131,** 3147-3156.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature biotechnology **19,** 242-247.

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis **16,** 1090-1094.

Wehrman, T., Kleaveland, B., Her, J.H., Balint, R.F., and Blau, H.M. (2002). Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. Proceedings of the National Academy of Sciences of the United States of America **99,** 3469-3474.

Weingartner, M., Criqui, M.C., Meszaros, T., Binarova, P., Schmit, A.C., Helfer, A., Derevier, A., Erhardt, M., Bogre, L., and Genschik, P. (2004). Expression of a nondegradable cyclin B1 affects plant development and leads to endomitosis by inhibiting the formation of a phragmoplast. The Plant cell **16,** 643-657.

Weinl, C., Marquardt, S., Kuijt, S.J., Nowack, M.K., Jakoby, M.J., Hulskamp, M., and Schnittger, A. (2005). Novel functions of plant cyclin-dependent kinase inhibitors, ICK1/KRP1, can act non-cell-autonomously and inhibit entry into mitosis. The Plant cell **17,** 1704-1722.

Whitford, R., Fernandez, A., De Groodt, R., Ortega, E., and Hilson, P. (2008). Plant CLE peptides from two distinct functional classes synergistically induce division of vascular cells. Proc Natl Acad Sci U S A.

Wildwater, M., Campilho, A., Perez-Perez, J.M., Heidstra, R., Blilou, I., Korthout, H., Chatterjee, J., Mariconti, L., Gruissem, W., and Scheres, B. (2005). The RETINOBLASTOMA-RELATED gene regulates stem cell maintenance in Arabidopsis roots. Cell **123,** 1337-1349.

Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A **98,** 3750-3755.

Witte, C.P., Noel, L.D., Gielbert, J., Parker, J.E., and Romeis, T. (2004). Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope. Plant molecular biology **55,** 135-147.

Wittig, I., Braun, H.P., and Schagger, H. (2006). Blue native PAGE. Nat Protoc **1,** 418-428.

Wolkowicz, R., Jager, G.C., and Nolan, G.P. (2005). A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. J Biol Chem **280,** 15195-15201.

Xing, D.H., and Chen, Z.X. (2006). Effects of mutations and constitutive overexpression of EDS1 and PAD4 on plant resistance to different types of microbial pathogens. Plant Sci **171,** 251-262.

Xu, X., Soutto, M., Xie, Q., Servick, S., Subramanian, C., von Arnim, A.G., and Johnson, C.H. (2007). Imaging protein interactions with bioluminescence resonance energy transfer (BRET) in plant and mammalian cells and tissues. Proceedings of the National Academy of Sciences of the United States of America **104,** 10264-10269.

Yamaguchi, M., Kato, H., Yoshida, S., Yamamura, S., Uchimiya, H., and Umeda, M. (2003). Control of in vitro organogenesis by cyclin-dependent kinase activities in plants. Proc Natl Acad Sci U S A **100,** 8019-8023.

Yan, X., Habedanck, R., and Nigg, E.A. (2006). A complex of two centrosomal proteins, CAP350 and FOP, cooperates with EB1 in microtubule anchoring. Mol Biol Cell **17,** 634-644.

Yang, P., Fu, H., Walker, J., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D. (2004). Purification of the Arabidopsis 26 S proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. J Biol Chem **279,** 6401-6413.

Yang, S., Yang, H., Grisafi, P., Sanchatjate, S., Fink, G.R., Sun, Q., and Hua, J. (2006). The BON/CPN gene family represses cell death and promotes cell growth in Arabidopsis. Plant J **45,** 166-179.

Yin, X.J., Volk, S., Ljung, K., Mehlmer, N., Dolezal, K., Ditengou, F., Hanano, S., Davis, S.J., Schmelzer, E., Sandberg, G., Teige, M., Palme, K., Pickart, C., and Bachmair, A. (2007). Ubiquitin lysine 63 chain forming ligases regulate apical dominance in Arabidopsis. The Plant cell **19,** 1898-1911.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.F., Dricot, A., Vazquez, A., Murray, R.R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.S., Motyl, A., Hudson, M.E., Park, J., Xin, X., Cusick, M.E., Moore, T., Boone, C., Snyder, M., Roth, F.P., Barabasi, A.L., Tavernier, J., Hill, D.E., and Vidal, M. (2008). High-quality binary protein interaction map of the yeast interactome network. Science **322,** 104-110.

Yu, Y., Steinmetz, A., Meyer, D., Brown, S., and Shen, W.H. (2003). The tobacco A-type cyclin, Nicta;CYCA3;2, at the nexus of cell division and differentiation. The Plant cell **15,** 2763-2777.

Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A., and Greenblatt, J.F. (2004). Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. J Proteome Res **3,** 463-468.

Zentella, R., Zhang, Z.L., Park, M., Thomas, S.G., Endo, A., Murase, K., Fleet, C.M., Jikumaru, Y., Nambara, E., Kamiya, Y., and Sun, T.P. (2007). Global analysis of della direct targets in early gibberellin signaling in Arabidopsis. The Plant cell **19,** 3037-3057.

Zhang, S., Cai, M., Xu, S., Chen, S., Chen, X., Chen, C., and Gu, J. (2002). Interaction of p58(PITSLRE), a G2/M-specific protein kinase, with cyclin D3. J Biol Chem **277,** 35314-35322.

Zhao, Q., Brkljacic, J., and Meier, I. (2008). Two Distinct Interacting Classes of Nuclear Envelope-Associated Coiled-Coil Proteins Are Required for the Tissue-Specific Nuclear Envelope Targeting of Arabidopsis RanGAP. The Plant cell.

Zhou, H., Watts, J.D., and Aebersold, R. (2001). A systematic approach to the analysis of protein phosphorylation. Nature biotechnology **19,** 375-378.

Zhou, Y., Wang, H., Gilmer, S., Whitwill, S., and Fowke, L.C. (2003). Effects of co-expressing the plant CDK inhibitor ICK1 and D-type cyclin genes on plant growth, cell size and ploidy in Arabidopsis thaliana. Planta **216,** 604-613.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., and Snyder, M. (2001). Global analysis of protein activities using proteome chips. Science (New York, N.Y **293,** 2101-2105.

Zimmermann, P., Hennig, L., and Gruissem, W. (2005). Gene-expression analysis and network discovery using Genevestigator. Trends in plant science **10,** 407-409.