Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana

Jelle Van Leene1,2, Jens Hollunder1,2, Dominique Eeckhout1,2, Geert Persiau1,2, Eveline Van De Sijlque1,2, Hilde Stals1,2, Gert Van Isterdael1,2, Aurine Verkest1,2, Sandy Neirynck1,2, Yelle Buffel1,2, Stefanie De Bodt1,2, Steven Maere1,2, Kris Laukens3, Anne Pharazyn4, Paulo CG Ferreira5, Nubia Eloy1,2,5, Charlotte Renne6, Christian Meyer6, Jean-Denis Faure6, Jens Steinbrenner7, Jim Beynon7, John C Larkin8, Yves Van de Peer1,2, Pierre Hilson1,2, Martin Kuiper1,2,10, Lieven De Veylder1,2, Harry Van Onckelen4, Dirk Inzé1,2, Erwin Witters4,9,11 and Geert De Jaeger1,2,11,*

1 Department of Plant Systems Biology, VIB, Ghent, Belgium, 2 Department of Plant Biotechnology and Genetics, Ghent University, Ghent, Belgium, 3 Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium, 4 Department of Biology, Center for Proteome Analysis and Mass Spectrometry, University of Antwerp, Antwerp, Belgium, 5 Instituto de Bioquimica Medicina, Centro de Ciencias da Saude, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 6 Institut Jean-Pierre Bourgin, INRA-AgroParisTech, Versailles Cedex, France, 7 School of Life Sciences, Warwick University, Warwick, UK, 8 Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA and 9 Flemish Institute for Technological Research (VITO), Mol, Belgium 10 Present address: Department of Biology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway 11 These authors contributed equally to this work

* Corresponding author. Department of Plant Systems Biology, VIB, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. Tel.: +32 9 3313870; Fax: +32 9 3313809; E-mail: geert.dejaeger@psb.vib-ugent.be

Received 20.11.09; accepted 12.6.10

Cell proliferation is the main driving force for plant growth. Although genome sequence analysis revealed a high number of cell cycle genes in plants, little is known about the molecular complexes steering cell division. In a targeted proteomics approach, we mapped the core complex machinery at the heart of the Arabidopsis thaliana cell cycle control. Besides a central regulatory network of core complexes, we distinguished a peripheral network that links the core machinery to up- and downstream pathways. Over 100 new candidate cell cycle proteins were predicted and an in-depth biological interpretation demonstrated the hypothesis-generating power of the interaction data. The data set provided a comprehensive view on heterodimeric cyclin-dependent kinase (CDK)–cyclin complexes in plants. For the first time, inhibitory proteins of plant-specific B-type CDKs were discovered and the anaphase-promoting complex was characterized and extended. Important conclusions were that mitotic A- and B-type cyclins form complexes with the plant-specific B-type CDKs and not with CDKA;1, and that D-type cyclins and S-phase-specific A-type cyclins seem to be associated exclusively with CDKA;1. Furthermore, we could show that plants have evolved a combinatorial toolkit consisting of at least 92 different CDK–cyclin complex variants, which strongly underscores the functional diversification among the large family of cyclins and reflects the pivotal role of cell cycle regulation in the developmental plasticity of plants.

Molecular Systems Biology 6:397; published online 10 August 2010; doi:10.1038/msb.2010.53

Subject Categories: plant biology; metabolic & regulatory networks

Keywords: Arabidopsis thaliana; cell cycle; interactome; protein complex; protein interactions

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar license to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

Introduction

The basic underlying mechanisms of cell division are conserved among all eukaryotes. However, the Arabidopsis thaliana (Arabidopsis) genome contains a collection of cell cycle regulatory genes (Vandepoele et al, 2002; Menges et al, 2005), which is intriguingly large when compared to other eukaryotes. In five regulatory classes, 71 genes are found in Arabidopsis versus only 15 in yeast and 23 in human (Supplementary Table 1). They encode cyclin-dependent kinases (CDKs), of which the substrate specificity is determined by association with various cyclins, whereas series of CDK activators and inhibitors regulate their activity (Inzé and De Veylder, 2006; De Veylder et al, 2007). Together with genes encoding the retinoblastoma-related (RBR) protein and members of the E2F/DP family, the genes for CDKs, cyclins, and their regulators were defined as the ‘core’ cell cycle genes in Arabidopsis (Vandepoele et al, 2002; Menges et al, 2005). This inventory was augmented with the discovery of genes involved in DNA replication (Shultz et al, 2007), and mitotic...
checkpoint homologs, including proteins of the anaphase-promoting complex (APC), an E3 ubiquitin ligase, which targets cell cycle proteins for degradation by the 26S proteasome (Capron et al., 2003). Microarray analysis demonstrated that many of these genes showed a cell cycle phase-dependent expression profile (Menges et al., 2005), whereas genetic studies confirmed their role in cell division (Inzé and De Veylder, 2006; De Veylder et al., 2007).

Despite the discovery of numerous cell cycle genes, little is known about the corresponding protein interaction network. Therefore, we applied tandem affinity purification (TAP) approach with the aim to isolate and analyze protein complexes for approximately 100 cell cycle proteins, of which most belong to the cell cycle core list (Supplementary Table II). As we focus on cell division and because plants contain only a minor fraction of dividing cells, we previously developed a TAP approach for complex isolation from Arabidopsis cell suspension cultures (Van Leene et al., 2007, 2008). These cell suspension cultures consist of undifferentiated dividing cells and therefore they not only serve as a model for plant meristems, but also are well suited to study protein interactions in the absence of developmental processes, pinpointing the basic cell cycle machinery (Menges et al., 2003). Furthermore, they provide an unlimited and cheap supply of proliferating cells that express more than 85% of the predicted core cell cycle genes. The expression of almost all core cell cycle regulators and related genes in cell suspension cultures is in agreement with the observation that most of them do not show strong tissue specificity (Menges et al., 2005). This approach allowed us to successfully map a first draft of the basic cell cycle complex machinery of Arabidopsis, providing many new insights into plant cell division.

Results and discussion

Mapping the cell cycle interactome

From the list of cell cycle genes described above, 102 proteins were selected as baits (Supplementary Table III). In addition, six interesting proteins that copurified with the baits were chosen for reverse TAP experiments. Cell cultures were stably transformed with transgenes encoding the tagged proteins under control of a constitutive promoter, as it had been previously shown that constitutive bait expression leads to higher complex recovery as compared to expression with endogenous promoters (Van Leene et al., 2007). Despite the use of this constitutive promoter that could induce artificial interactions, we observed that accumulation levels of the fusion proteins depend to a large extent on the nature of the bait and are not always higher than those of the corresponding endogenous protein (Van Leene et al., 2007). A plausible explanation is the high level of posttranslational regulation among many essential cell cycle proteins. Moreover, given the high ploidy level (8n) of the Arabidopsis cultures we used, the average transgene copy number per cell might be lower than that of the corresponding endogenous gene. A major advantage of this constitutive promoter was that 95% of the baits were successfully produced as TAP-tagged fusion proteins (Figure 1A). At least two independent purifications were performed for each of the expressed baits on extracts from non-synchronized exponentially growing cells, for a total of 303 purifications. A flow cytometric analysis of the collected cell material showed an equal G1–G2 phase distribution (Van Leene et al., 2007), covering all cell cycle phases, meaning that apparently interacting proteins might represent different alternative cell phase-specific complexes. Purified proteins were separated on gel and identified by tandem matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS). Nonspecific interactors and background proteins were determined through control purifications on wild-type cell culture extracts (mock) or extracts from cultures expressing tagged fusions of heterologous green fluorescent protein (GFP), red fluorescent protein (RFP) or β-glucuronidase (GUS; Supplementary Table IV). All proteins identified in these control experiments were subtracted systematically. Although most false positive interactions are discarded by this approach, some artificial and bait-specific interactions might remain in the data set, for example, interactions occasionally generated during cell lysis. Next, redundancy was filtered out for reciprocal interactions found in both directions, retaining the interaction with the highest MS scores. As such, a final non-redundant data set of 857 interactions among 393 proteins was obtained. As a first proof of the robustness of the data, all six reverse TAP experiments confirmed the original interaction. To further assess the quality of the data set and to evaluate its novelty, we screened for the overlap with protein–protein interactions present in public databases and observed that 82% of the interactions are not yet documented in TAIR (Swarbreck et al., 2008), InTact (Kerrien et al., 2007), Arabidopsis Reactome (Tesmetsiz et al., 2008), AtPID (Cui et al., 2008), Reactome (Vastrik et al., 2007) and The Bio-Array Resource (BAR) for Arabidopsis Functional Genomics (Geisler-Lee et al., 2007), providing a huge amount of new information. In addition, this analysis demonstrates the reliability of the data set because 150 known or predicted interactions were confirmed.

We visualized our data set as a network graph according to the ‘spoke’ model in which proteins share co-complex membership with their immediate interactors through direct or indirect physical binding (Supplementary Figure 1). Hence, caution is required for the interpretation of the data, in particular because bait–prey interactions identified in TAP experiments might not actually represent direct physical interactions. Additional information on the proteins (e.g. periodicity during the cell cycle, localization) or on the bait–prey relationships (e.g. degree of co-expression, known or new interaction) was integrated into the interactome (Figure 2; see also Supplementary Tables V and VI). Moreover, the entire interactome and all discussed subnetworks can be easily visualized through a Cytoscape (Shannon et al., 2003) web start, and the data can be consulted in a matrix pivot table (see Figure 3 for more explanation). All protein interactions have been submitted to the IMEx (http://imex.sf.net) consortium through IntAct (Kerrien et al., 2007) with the assigned identifier IM-9598.

Computational quality assessment

To characterize the interactome, we compared two domains that had been identified in the global data set: domain I,
containing 371 interactions confirmed in at least two independent experimental repeats or in the reciprocal purification experiment and domain I2 consisting of 486 uniquely observed interactions (Figure 1A; Supplementary Figure 1). Several observations underlined the quality of both domains. First, the merged interactions between baits formed a single highly interconnected network (Supplementary Figure 2), reflecting their common involvement in the same biological process. Second, genes encoding preys in domain I1 and I2 are enriched among sequences with E2F and M-specific activator (MSA) promoter elements, involved in the G1-to-S and G2-to-M transition, respectively (Figure 1B), thereby demonstrating the successful purification of complexes functioning at S-phase and mitosis. Third, the transcript Pearson Correlation Coefficients (PCCs) that reflect the degree of co-expression correlation were calculated for all interactions, based on an Arabidopsis ATH1 micro-array compendium of experiments focusing on cell cycle or plant growth and development (Supplementary Table VII). On average, a transcript PCC of 0.324 was found for interactions of domain I1 and 0.144 for domain I2, which is significantly higher than that of the average PCC of 100 corresponding random networks (0.016; Figure 1C). These PCCs can be used for confidence assignment to new interactions, because interactors with strong expression correlation are often part of a common molecular assembly (Gunsalus et al., 2005). Finally, we identified a large number of new candidate cell cycle genes among the preys of both domains. Therefore, we integrated different cell cycle-related features (Supplementary Table VIII), including periodicity during cell division, CDK phosphorylation sites, and cell cycle-related promoter and protein destruction motifs. In a set of 518 known cell cycle genes (Supplementary Table IX), compiled based on gene ontology (GO) annotation, and supplemented with genes involved in the cell cycle (Vandepoele et al., 2002; Capron et al., 2003; Menges et al., 2005; Shultz et al., 2007), a clear enrichment compared to the whole gene pool was detected for genes possessing more than one cell cycle feature (Supplementary Figure 3). The same was true for our bait list, validating the choice of the baits, and for the domain I1 and I2 prey lists after subtracting known cell cycle proteins (Figure 1D). Finally, 40 new candidates with more than one feature were extracted from domain I1 and 83
from domain I₂ (Supplementary Table X), yielding a total of 106 new candidate cell cycle proteins.

Besides common qualities of both interactome domains, their real significance appeared through mutual differences. For example, 51% of the interactions in domain I₁ are between bait proteins, down to only 9% in domain I₂. Preys periodically expressed during the cell cycle were enriched solely in domain I₁ (Figure 1B), and although both prey sets were enriched for the GO term ‘cell cycle’, this was not the top hit in domain I₂. Here, and in contrast to domain I₁, significant enrichment was observed for GO categories related to stress response, phytohormone stimuli, or energy derivation (Supplementary Table XI). A GO similarity analysis (De Bodt et al., 2009) between pairs of bait and prey confirmed that, in general, pairs of domain I₂ shared only cellular component, a prerequisite for interacting proteins, but not biological process or molecular function, in contrast to pairs of domain I₁ (Supplementary Figure 4). These observations reveal two subspaces in the cell cycle interactome: a central regulatory network of stable complexes that are repeatedly isolated and represent core regulatory units, and a peripheral network comprising transient interactions identified less frequently, which are involved in other aspects of the process, such as crosstalk between core complexes or connections with other pathways. Additional evidence for the difference between these two subspaces was obtained when the preys were screened for the presence of CDK consensus phosphorylation sites. An enrichment for preys with CDK phosphorylation motifs was found to be statistically significant only in domain I₂ (Figure 1B), but not in domain I₁, nor among the uniquely observed interactions of a TAP data set generated with 54 baits unrelated to cell cycle mechanisms (data not shown). Moreover, this observation emphasizes the regulatory role of protein phosphorylation by CDK–cyclin complexes on proteins present in cell cycle-linked pathways.

**Biological validation in Arabidopsis plants**

As the TAP data set had been generated from complexes purified from cell suspension cultures, we further validated its biological relevance in Arabidopsis plants. For this purpose, we conducted a transient split-luciferase (LUC) assay in cotyledons of Arabidopsis seedlings (Supplementary Figure 5), in which the firefly LUC protein is split into two halves and fused to two different proteins. The firefly LUC activity is only reconstituted when the N- and C-terminal LUC moieties are brought together by the two interacting proteins. This interaction can be visualized with a low-light imaging system (Chen et al., 2008). Background signal was defined from untransformed seedlings, for which data were gathered from 27 different experiments providing a mean relative LUC activity of 535 (Untr.). We arbitrary set a relative LUC activity threshold of 4500 for a positive interaction, which is nine-fold higher than that from untransformed seedlings, meaning that values equal or above this threshold is assumed to represent a genuine interaction. A total of 17 new protein pairs were selected for the split-LUC analysis, covering eight interactions from domain I₁ and nine from domain I₂ (Figure 2; Supplementary Table XII). As negative controls, two protein pairs that did not interact in the TAP data set were tested in the split-LUC assay and these scored negative (Figure 2). Pairs 5, 6, and 7 are not expected to interact in the TAP data set because the light signals were beneath the threshold, although they might represent true, although weak, interactions because the light signals were considered negative because their relative LUC activity was less than 500, meaning that their expression values were below the threshold of 4500 for a positive LUC interaction.
showed that the overlap between binary interaction data and affinity-purified complex data is less than 20% in yeast, demonstrating that different protein interaction techniques detect different subspaces of the total interactome of an organism (Yu et al., 2008). Hence, bait–prey interactions discovered by TAP that were negative in this split-LUC assay might represent indirect interactions. Taken together, these validation experiments reveal that both domains contain a large portion of highly reliable interactions and that the interactions uncovered in cell suspension might well be extrapolated \textit{in planta}.

**A bird's eye view on the cell cycle interactome**

With respect to insights into the cell cycle physiology, the interactome was subdivided according to the functional domains of the cell cycle. The main parts of the interactome are represented in subnetworks (Fig. 3), which can be downloaded and explored through a Cytoscape web start at http://www.psb.ugent.be/supplementary-data-gejae/512-interactome (username: interactomics, password: CCinteractome).

### Figure 3

Subnetworks representing the main parts of the interactome. The subnetworks discussed in detail are (A) CDK–cyclin core complexes, (B) new interactions with CDK–cyclin complexes, (C) positive regulation by CAKs, (D) negative regulation by KRPs, (E) negative regulation by SIM and SIM-related proteins, (F) DNA replication complexes, (G) the anaphase-promoting complex, and (H) spindle checkpoint complexes. The degree of coexpression correlation (PCC) between a gene pair is given as an edge attribute in the color of the edge. Known and predicted interactions were obtained from public databases and can be distinguished from new interactions through the thickness of the edge. Information about which database documented the interaction is provided as an edge attribute in the Cytoscape file in the edge attribute browser. The edge style (solid versus dashed) reflects whether an interaction is confirmed in an experimental repeat or in the reciprocal purification (domain I1) or the interaction was uniquely observed (domain I2). Information about the proteins that were used as bait and about the newly predicted cell cycle proteins is integrated into the color of the nodes. Periodic genes, showing periodicity at transcriptional level during the cell cycle according to a gene list compiled as described in Materials and methods section, are marked with a blue border. The shape of the node refers to the cell cycle phase in which the gene expression peaks according to Menges et al. (2003). NA, not assessed. The entire interactome and all subnetworks are also available through a Cytoscape web start at http://www.psb.ugent.be/supplementary-data-gejae/512-interactome (username: interactomics, password: CCinteractome). Moreover, the data are presented at this location as an excel pivot table in matrix format allowing easy querying using baits and preys arranged on horizontal and vertical lines, respectively. The filter buttons of the bait allow easy filtering of its preys, and scrolling horizontally gives a clear idea of the specificity of the interaction or crosstalk with other complexes. The numbers in the matrix represent how many times a prey copurified with a given bait. The presence of cell cycle-related features among the preys (Supplementary Table VIII) is also implemented on the horizontal line at the end of the table. Source data is available for this figure at www.nature.com/msb.
classes of the baits (Figure 3) and core protein complexes were derived.

The core CDK–cyclin complexes

From the subnetwork between A- and B-type CDKs and their cyclins (Figure 3A), we extracted CDK–cyclin complexes. CDKA;1, encoded by the constitutively expressed ortholog of yeast cdc28, formed stable complexes with constitutively expressed D-type cyclins (e.g. CYCD2;1), with fluctuating D-type cyclins presumed to be active at the cell cycle entry, the G1-to-S transition, or the M-phase (e.g. CYCD3;1), and with S-phase-specific A3-type cyclins (Supplementary Table XIII). On the contrary, A2-type and all B-type cyclins, with an expression peak at the G2-to-M boundary, only bound the plant-specific mitotic B-type CDKs that regulate entry into and progression through mitosis (Menges et al., 2005). The B1-type CDKs, peaking earlier than B2-type (Menges et al., 2005), bound A2 and B2 cyclins, whereas B2-type CDKs were found exclusively with B1-type cyclins with transcript levels peaking late in M-phase. These results show that our data support transcriptome data (Menges et al., 2005) and give a comprehensive overview of A- and B-type CDK–cyclin complexes. Our data imply that mitotic A- and B-type cyclins exclusively form heterodimeric complexes with the plant-specific B-type CDKs and not with CDKA;1, whereas D-type cyclins associate with CDKA;1. In agreement with the previously described interaction of CDKB2;1 with CYCD4;1 (Kono et al., 2003), we observed that CDKB1;1 associated with CYCD4;1; however, as this interaction was found only once, it might reflect a more transient interaction, such as phosphorylation of CDKB1;1 by CDKA;1–CYCD4;1 complexes that regulate its activity.

All CDK–cyclin core complexes copurify at least one of the two scaffold CDK subunit (CKS) proteins. Additional interesting new interactions were identified with core CDKs and cyclins (Figure 3B). A protein of unknown function, AT4G14310, copurified with CDKA;1, CKS1, CKS2, CYCA3;1, CYCA3;4, and KRP2. Reverse purification confirmed interaction with CDKA;1 and CKS2, and revealed interaction with the plant-specific kinesin motor protein KCA2. As the latter is involved in division plane determination (Vanstraelen et al., 2006), AT4G14310 might be involved in the same pathway, as supported by the observation that a GFP fusion of AT4G14310 localized at the pre-prophase band (data not shown). Furthermore, CDKA;1 copurified with RPN1a, a regulatory subunit of the 26S proteasome complex, possibly reflecting cell cycle regulation of the 26S proteasome. In addition, 19 proteasome-related proteins were pulled down with this RPN1a subunit of the 26S proteasome. Further, CDKA;1 interacted with three proteins (one phosphoglucomutase and two UDP-glucose 6-
dehydrogenases) from the UDP-xylose biosynthesis pathway, coupling cell cycle regulation with cell wall synthesis (Seifert, 2004). The spindle pole body component 98 and γ-tubulin, two proteins involved in microtubule (MT) nucleation, were pulled down with CYCB1;3. These proteins colocalize at nuclear membranes during G2-phase and are involved in the assembly of pre-prophase band, a plant-specific structure required for polarity determination during cell cycle (Erhardt et al, 2002). The interaction with CYCB1;3 appears to be justified because activation of MT nucleation sites and coordinated regulation of the MT assembly might be controlled by cell cycle and/or developmental signals.

Positive regulation of CDK–cyclin complexes

We mapped complexes of CDK-activating kinases (CAKs) that activate CDKs through phosphorylation of a threonine residue in their T-loop. Arabidopsis encodes four CAKs, namely three D-type CDKs and one CAK-activating kinase CDKF;1. Both CDKD;2 and CDKD;3 copurify with CYCH;1 and the CAK assembly factor MAT1 (Figure 3C), which indicates the presence of trimeric CDKD–CYCH–MAT1 complexes in plants, similar to the mammalian D-type CDKs (Devault et al, 1995). As in rice (Oryza sativa; Rohila et al, 2006), CDKD;2 is also part of the basal TFIIH complex involved in transcription and DNA repair, of which three members (UVH6/XPD, AT1G55750 and AT4G17020) copurified. 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). In addition, the interaction of CDKD;2 with three ribose-phosphate pyrophosphokinases involved in nucleotide biosynthesis, further demonstrated the role of D-type CDKs in the S-phase. Besides the extraction of complexes previously shown in other organisms, our data also suggested new functional links. For example, CDKF;1 does not only interact with its known target CDKD;2 (Shimotohno et al, 2004), but also with CDKG;2 (Figure 3C). The G-type CDK class has two members in Arabidopsis and is homologous to the human p58 galactosyltransferase protein associated with cytokinesis (Menges et al, 2005). A cyclin with no clear function in the cell cycle, CYCL1, copurified with both CDKGs, validating the clustering of CYCL1 with CDKG;2 in a tissue-specific gene expression analysis (Menges et al, 2005). The other proteins forming complexes with CDKG (Figure 3C) and the SR-like splicing domain of CYCL1 (Forment et al, 2002) hint at a function in transcription and transcript processing. On the basis of the peak of CDKG transcription (Menges et al, 2005), CDKG–CYCL complexes are presumably active at the cell cycle onset.

Negative regulation of CDK–cyclin complexes

Plants not only contain CDK activators but also many negative regulators (Supplementary Table I and II), such as the seven Kip-related proteins (KRPCs) in Arabidopsis. The proteins KRP2–KRP7 copurified solely with CDKA;1 and D-type cyclins, suggesting that KRPCs only inhibit CYCD–CDKA;1 complexes (Figure 3D). In addition, KRPCs have been postulated to regulate the nuclear import of cell cycle regulators (Zhou et al, 2006). Their association with transcription factors suggests that their role in reallocation is not solely targeted to CDK–cyclin complexes (Figure 3D). Another family of cell cycle inhibitors, upregulated by biotic and abiotic stress, comprise SIAMESE (SIM) and SIAMESE-related (SMR) proteins (Churchman et al, 2006; Peres et al, 2007). Thus far, this family consisted of six proteins (SIM and SMR1–SMR5), but a recent analysis has extended this family with eight additional members (J Van Leene et al., unpublished data; Supplementary Tables I and II). Endoreduplication in trichomes is promoted by SIM by the suppression of mitosis, possibly through inhibition of CDKA;1–CYCD complexes (Churchman et al, 2006). However, in our analysis, CDKB1;1, and not CDKA;1 copurified with SIM as bait (Figure 3E), suggesting that endoreduplication might be triggered directly by inhibition of mitotic CDKB complexes. In addition, SMR1 and SMR2 associate with CDKB1;1, and its interactor CYCB2;4 binds AT2G28330 (SMR1), one of the additional members of the SMR family. Until now, such potential inhibitors of B-type CDKs had not been found in plants. In contrast, SMR3–SMR5 and two new members of the SMR clan, AT5G40460 (SMR6) and AT1G10690 (SMR8), clearly associate with CDKA;1 and D-type cyclins. The latter was confirmed by reverse purifications. As SMR6 was induced almost 20-fold in plants co-over-expressing E2Fa and DPa (Vandepoele et al, 2005), it might inhibit CDKA;1–CYCD complexes during S-phase, preventing the re-initiation of DNA replication. Similar to KRPCs, we also observed nuclear import proteins and transcription factor-related proteins with the SMRs.

Progression to DNA replication and through mitosis

At the G1-to-S boundary, CDK–cyclin complexes activate the E2F–DP pathway by phosphorylation of the repressor RBR, inducing genes involved in nucleotide synthesis, DNA replication, and DNA repair. We confirm that E2Fa and E2Fb associate both with DPa and DPb, and that all E2Fs, including E2Fc, and DP proteins interact with RBR (Figure 3F). Intriguingly, the mitotic CDKB1;1, and not CDKA;1, copurified with RBR, providing additional evidence that the E2F–DP–RBR network is not only active at G1-to-S, but also at G2-to-M transition, as previously suggested for plants (Magyar et al, 2005), Drosophila (Neufeld et al, 1998), and mammalian cells (Ishida et al, 2001). Furthermore, because CDKB1;1 interacted with DEL3, an atypical E2F factor lacking the trans-activating domain (Lammens et al, 2009), we propose that activity of DEL3 might be regulated by CDKB1;1 at the G2-to-M transition, consistent with the observation that both genes are transcribed at that time point and that the encoded proteins can indeed meet each other (Menges et al, 2005). Additional complexes involved in DNA replication or repair were isolated (Figure 3F), such as the minichromosome maintenance (MCM) complex, a complex containing the proliferating cell nuclear antigen 1 (PCNA1), which is a sliding clamp for DNA polymerase, the alternative Ctf18 replication factor C complex required for sister chromatid cohesion in yeast (Mayer et al,
and a complex involved in stabilization of single-stranded DNA during replication, repair and transcription, including RPA2, two RPA3 proteins, and a putative replication protein (AT2G06510; Shultz et al., 2007). The MCM and PCNA complexes are clearly enriched for genes with an E2F motif in their promoter (Supplementary Figure 6), meaning that these complexes are synthesized and assembled just in time at the beginning of S-phase through the E2F–DP pathway. The presence of E2F motifs in the promoters of E2Fb and DPa or E2Fc and RBR reflects positive or negative feedback mechanisms regulating this pathway, respectively (Vandepoele et al., 2005).

From the G2-to-M transition onward, unidirectional progression through the cell cycle is, next through the action of CDK–cyclin complexes, further achieved by the APC complex that targets cell cycle proteins for destruction by the 26S proteasome (De Veylder et al., 2007). For the first time, a plant APC has been isolated biochemically, and the complex is visualized as a very tightly interconnected network enriched for highly co-expressed gene pairs (Figure 3G). All putative plant APC subunits were identified, except the two small proteins (<10 kDa) APC13 (Bonsai) and Cdc26. However, these proteins were shown not to be essential for proper APC functioning in yeast (Thornton and Toczyski, 2006) and therefore might not belong to the active core complex in plants. We further demonstrate that both Cdc27a and Cdc27b (HOBBIT) can be part of the APC. Three new plant-specific APC interactors (AT1G32310, UV14, and UV14-like) were identified and their interaction with the APC was confirmed by reverse purification. The interactor UV14 has been postulated to keep cells in the mitotic state because mutants for UV14 showed increased endoreduplication (Hase et al., 2006). It is intriguing that this function can be linked now with the APC. A closer look at the protein sequence of UV14 and UV14-like revealed different CDK consensus phosphorylation motifs in their sequences, possibly important in the regulation of their activity. Interestingly, both proteins have a C-terminal methionine–arginine (MR) tail. This MR tail is present in only 40 Arabidopsis proteins, whereas in Xenopus it is implicated 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 (Vodermann et al., 2003). Regulation of APC activity could be achieved by CDKA1, as derived from its interaction with APC10. Analogous to yeast, APC activators in plants are most probably guided to the APC by the action of the CCT chaperonin (Camasses et al., 2003), because three such family members copurified.

This chaperonin could also assist in the assembly of other spindle checkpoint complexes, as shown for three Arabidopsis homologs of the mitotic checkpoint proteins Mad2 and Bub3 (Capron et al., 2003; Menges et al., 2005; Figure 3H). Unattached kinetochores trigger the formation of Mad2–Bub3–BubR1 complex that, in turn, inhibits Cdc20 APC activators, thereby preventing degradation of several cell cycle regulators and progression of anaphase (Kimbara et al., 2004). The mitotic checkpoint proteins pulled down many specific interactors, including M-phase-specific kinases; the highly co-expressed MAP65-3, located at mitotic microtubule arrays and essential for cytokinesis (Müller et al., 2004); histone H4; two peptidyl-prolyl cis-trans isomerase; two proteins of the prefoldin chaperone; a helicase (AT1G24290) similar to the replication factor C protein; and an ADP-ribosylation factor GTase-activating protein (AT3G15970). The latter two proteins had previously been predicted to interact with MAD2-like (Geisler-Lee et al., 2007).

The extraction of the subnetwork of genes with an MSA motif in their promoter revealed two small clusters (Supplementary Figure 7), representing a module of the APC connected to a module of a mitotic checkpoint complex by an unknown protein (AT4G28230). Interestingly, DPa and E2Fb possess an MSA motif, supporting our previous hypothesis that the E2F–DP–RB pathway is also active at the G2-to-M transition.

**An integrative view on CDK–cyclin complexes**

It has been shown that transcriptional diversification during cell cycle progression is a key element among cyclins and core cell cycle regulators in plant cells (Menges et al., 2005). We ranked each cyclin along the cell cycle phases according to their peak during transcription and grouped them with their associated CDKs, CKS scaffolding proteins, and negative regulators (Figure 4; Supplementary Table XIII). Modules of interacting proteins were obtained showing an assorted set of CDK–cyclin complexes with high regulatory differentiation. Even within the same subfamily (e.g. cyclin A3, B1, B2, D3, and D4), cyclins differ not only in their functional time frame but also in the type and number of CDKs, inhibitors and scaffolding proteins they bind, further indicating their functional diversification. According to our interaction data, at least 92 different variants of CDK–cyclin complexes are found in Arabidopsis. We speculate that further complex analysis in synchronized cultures will demonstrate an even higher variety. In conclusion, these results reflect how several rounds of gene duplication (Sterck et al., 2007) allowed the evolution of a large set of cyclin paralogs and a myriad of regulators, resulting in a significant jump in the complexity of the cell cycle machinery that could accommodate unique plant-specific features, such as an indeterminate mode of postembryonic development. Through their extensive regulation and connection with multiple up- and downstream pathways, the core cell cycle complexes might offer the sessile plant a flexible toolkit to fine-tune cell proliferation in response to an ever-changing environment.

**Materials and methods**

Cloning of transgenes encoding tag fusions under control of the constitutive cauliflower mosaic virus 35S promoter, transformation of Arabidopsis cell suspension cultures, protein extract preparation, TAP purification, protein precipitation, and separation were carried out as previously described (Van Leene et al., 2007). The adapted protocol used for purification of protein complexes incorporating GS-tagged (Bürcsküster et al., 2006) bait has been described previously (Van Leene et al., 2008). For identification by MS, minor adjustments were implemented compared to previously described protocols (Van Leene et al., 2007), as described below.
Proteolysis and peptide isolation

After destaining, gel slabs were washed for 1 h in H2O, polypeptide disulfide bridges were reduced for 40 min in 25 ml of 6.66 mM DTT in 50 mM NH4HCO3 and the thiol groups were alkylated sequentially for 30 min in 55 mM iodacetamide in 50 mM NH4HCO3. After washing the gel slabs three times with water, complete lanes from the 50 mM NH4HCO3 and 10% CH3CN (v/v) 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 tandem MS instrument (4700 and 4800 Proteomics Analyzer; Applied Biosystems) was used to acquire peptide mass fingerprints (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 ± 10 p.p.m. for each analyzed peptide spot on the MALDI target. Using the individual PMF spectra, up to 16 peptides, exceeding a signal-to-noise ratio of 20, which passed through a mass exclusion filter, were submitted to fragmentation analysis.

MS-based protein homology identification

Data search files were generated with the search engine settings presented previously (Van Leene et al., 2007; Supplementary MS data) and submitted for protein homology identification against the TAIR 8.0 database (Swarbreck et al, 2008) by using a local database search engine (Mascot 2.1, Matrix Science). Protein homology identifications of top hits (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. Preferentially, identifications resulting from combined searches, PMF, and MS/MS were retained. In addition, PMF-only identifications were retained, but an additional restriction was implemented here. To reduce the number of false positive identifications, PMF-only identifications, for which 50% or more of the matched peptides had a trypsin miscleavage, were discarded.
Data analysis

Enrichment analyses

For the periodic gene identification and enrichment analysis, a list of 1258 genes showing cell cycle-regulated and cell cycle-associated expression was compiled from two data sets (Menges et al., 2003; Jensen et al., 2006). Genome wide corresponds to all 23,834 genes present on the Affymetrix ATH1 microarray.

Analysis of overrepresentation of GO terms was done using 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 (Benjamini and Yekutieli, 2001). The Arabidopsis gene annotation file used in the analysis was downloaded from the GO website on the October 4, 2008.

Genes containing E2F or MSA motifs in their promoter sequence were in silico determined by combining transcript expression data and comparative genomics (Vandepoele et al., 2006). Here, genome wide corresponds to 19,173 genes for which a unique probe set is available on the ATH1 microarray.

Proteins containing the CDK consensus phosphorylation site [ST]P[X][KR], a known hallmark of CDK substrates (De Veylder et al., 1997), were considered as potential CDK substrates. The presence of the consensus motif was screened with the patmatch tool available at TAIR (Swarbreck et al., 2008), and hence, genome wide corresponds here to all 27,235 proteins present in the TAIR8.0 release.

All enrichment analyses were compared to the genome-wide situation and P-values were calculated using a hypergeometric cumulative distribution function. Proteins that could not be assigned to a specific gene locus were discarded from all enrichment analyses.

GO similarity analysis

To calculate the GO similarity scores, GO terms were extracted from the GO database (Ashburner et al., 2000) and annotations for Arabidopsis proteins were downloaded from TAIR (Swarbreck et al., 2008). For each protein pair, all GO terms of both proteins were compared to each other and GO similarity scores were calculated as described previously (De Bodt et al., 2009). 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:00003673), is calculated. Subsequently, the maximum value of the calculated depths is taken as the GO similarity score for a certain protein pair. The assignment of GO terms based on physical interactions (IPI) or electronically assigned and less reliably assigned GO terms (with evidence codes ND, NR, NAS, and IEA) were removed. In addition, GO similarity scores were calculated for gene pairs from 1000 randomized data sets. For the comparison original versus random network, we considered the subnetwork of the protein interaction network in which each protein–protein interaction could be overlaid with at least one GO term. On the basis of this original subnetwork, we generated 1000 random networks maintaining the number of nodes and interactions containing at least one GO term between the nodes. The nodes were randomly selected from a pool containing all Arabidopsis proteins with at least one GO term. The networks of the three GO categories: biological process, molecular function, and cellular component were analyzed.

Co-expression analysis

Transcript PCCs reflecting the degree of co-expression correlation were calculated on the basis of an Arabidopsis ATH1 microarray compendium of 518 experiments focused toward plant growth and development (Supplementary Table VII). We compared the PCC distribution of both data sets with the PCC distribution of 100 randomized data sets. Subsequently to the GO similarity analysis, we considered a subnetwork of our protein interaction network containing those interactions that could be addressed with an expression value based on our compendium. The nodes of the random networks were randomly selected from a pool of all Arabidopsis proteins, whereas the number of expression links was maintained.

New candidate cell cycle proteins

Periodic genes, genes with E2Fa-like or MSA-like motifs in the promoter sequence and proteins containing a CDK consensus motif were determined as described above (enrichment analyses). All genes containing the remaining cell cycle-related promoter motifs were determined by the same in silico analysis combining transcript expression data and comparative genomics (Vandepoele et al., 2006). Proteins containing a PEST motif were determined using the pepstain tool provided by EMBoss (Rice et al., 2000). The D-box corresponds to the amino-acid motif RxxLxxxxN where x is any amino acid, whereas the A-box corresponds to the amino-acid motif QRVL. The GxEN-box corresponds to the amino-acid motif GxEN where x is any amino acid, whereas the KEN-box corresponds to the amino-acid motif KEN. All enrichment analyses were compared to the genome-wide situation and P-values were calculated using a hypergeometric cumulative distribution. Proteins that could not be assigned to a specific gene locus were discarded from all enrichment analyses.

Generation of the network in Cytoscape

The cytoscape file, which is accessible through a Cytoscape webstart, was generated in Cytoscape 2.5.1. by importing node and edge attribute files (Supplementary Tables V and VI) representing information regarding proteins and interactions, respectively.

Transient split-luciferase analysis

Firefly luciferase split constructs were generated as described previously (Chen et al., 2008). Sequences coding for the N- or C-terminal luciferase moieties were cloned in Gateway pPZP200-based vectors (Karimi et al., 2007), either 5′ (for N-terminal fusions of luciferase) or 3′ (C-terminal fusions) to the gateway recombination sequence. This way, we obtained four different vector combinations, allowing insertion of the gene of interest through gateway recombination cloning. Transient transformation was carried out on 15–20 seedlings of Landsberg erecta Arabidopsis seedlings grown in six-well plates as described previously (Marion et al., 2008). Transformed seedlings were sprayed twice with a 5 mM luciferine solution (synchem OHG) containing 0.01 % Triton X-100 and imaged with an ultra-amplified CCD camera (Photonic Science). Levels of light emissions were obtained after integrating 2000 images (Photolite 32 software). Light signals were quantified within a region of interest corresponding to the entire well and background was subtracted to obtain net light emission. Finally, luciferase emission was normalized according to the number of infiltrated plants estimated by their autofluorescence values. Leaf autofluorescence was measured with an excitation at 635 nm and a long pass filter $+665$ nm (FLA5000,FUJII).

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (http://www.nature.com/msb/).

Acknowledgements

We thank Martine De Cock for help in preparing the paper. This work was supported by grants from the Institute for the Promotion of Innovation through Science and Technology in Flanders (GBOU grant no. 20193), the Interuniversity Attraction Poles Programme (IUAP V1/33 and V1/25), initiated by the Belgian State, Science Policy Office (BELSPO), the European Commission (AGRON-OMICS; LSHG-CT-2006-037704), and the National Science Foundation (grant IOS 0744566 to JCL). AV thanks Agency for Innovation by Science and Technology in Flanders for a postdoctoral fellowship. NE was the recipient of a BELSPO postdoctoral fellowship. JVL, SDB, and SM are postdoctoral fellows of the Research Foundation-Flanders.

Conflict of interest

The authors declare that they have no conflict of interest.
References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25: 25–29

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Statist 29: 1165–1188

Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superici-Furiga C, Bauch A (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat Methods 3: 1013–1019

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

Capron A, Krézé L, Genschik P (2003) First glance at the plant APC/C, a highly conserved ubiquitin-protein ligase. Trends Plant Sci 8: 83–89

Chen H, Zou Y, Shang Y, Lin H, Wang Y, Cai R, Tang X, Zhou J-M (2008) Firefly luciferase complementation imaging assay for protein–protein interactions in plants. Plant Physiol 146: 368–376

Churchman ML, Brown ML, Kato N, Kirik V, Hülskamp M, Inze D, Hase Y, Trung KH, Matsunaga T, Tanaka A (2006) A mutation in the D999–D1008

Churchman J, Larkin JC (2006) SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana. Plant Cell 18: 3145–3157

Cui J, Li P, Li G, Xu F, Zhao C, Li Y, Yang Z, Wang G, Yu Q, Li Y, Shi T (2008) AtPID: Arabidopsis thaliana protein interactome database—an integrative platform for plant systems biology. Nucleic Acids Res 36: D999–D1008

De Bodt S, Proost S, Vandepoele K, Rouze P, Van de Peer Y (2009) Predicting protein–protein interactions in Arabidopsis thaliana through integration of orthology, gene ontology and co-expression. BMC Genomics 10: 288

De Veylder L, Segers D, Glab N, Van Montagu M, Inze D (1997) Identification of proteins interacting with the plant cell cycle regulator Cdc20. Proc Natl Acad Sci USA 94: 317–326

De Veylder L, Walker JD, Zheng Z, Oppenheimer DG, Gwin T, Churchman J, Larkin JC (2006) SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana. Plant Cell 18: 3145–3157

Erhardt M, Stoppin-Mellet V, Campagne S, Canaday J, Mutterer J, Fabian T, Sauter M, Muller T, Peter C, Lambert A-M, Schmit A-C (2002) The role of the Arabidopsis E2F2 transcription factor family. Trends Cell Biol 12: 111–118

Gaasterland E, Heymans K, Kuiper M (2005) BiNGO: a cytoskeleton plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448–3449

Hayes MJ, Kimata Y, Wattam SL, Lindon C, Mao G, Yamanishi H, Fromm ME (2006) The role of the Arabidopsis E2F2 transcription factor family. Trends Cell Biol 12: 111–118

Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, Bork P (2006) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504

Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, Bork P (2006) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504

Kimbara J, Endo TR, Nasuda S (2004) Characterization of the genes encoding for MAD2 homologues in wheat. Chromosoma Res 12: 703–714

Kono A, Umeda-Hara C, Lee J, Ito M, Uchimiya H, 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

Lammens T, Li J, Leone G, De Veylder L (2009) Atypical E2Fs: new players in the E2F transcription factor family. Trends Cell Biol 19: 11–18

Maire S, Heymans K, Kuiper M (2005) BiNGO: a cytoskeleton plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448–3449

Menges M, de Jager SM, Gruissem W, Murray JAH (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

Menges M, Hennig L, Gruissem W, Murray JA (2003) Genome-wide gene expression in an Arabidopsis cell suspension. Plant Mol Biol 53: 423–442

Müller S, Smetenko A, Wagner V, Heinrich M, Hussey PJ, Hauser M-T (2004) The plant microtubule-associated protein AtMAP65-3/PLE is essential for cytoskeletal phragmoplast function. Curr Biol 14: 412–417

Neufeld TP, de la Cruz AFA, Johnston LA, Edgar BA (1998) Coordination of growth and cell division in the Drosohila wing. Cell 93: 1183–1193

Peres A, Churchman ML, Hariharan S, Himanan K, Verkest A, Vandepoele K, Magyar Z, Hatzfeld Y, Van Der Schueren E, Beemer GET, Frankard V, Larkin JC, Inze D, De Veylder L (2007) Novel plant-specific cyclin-dependent kinase inhibitors induced by biotic and abiotic stresses. J Biol Chem 282: 25588–25596

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

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

Seifert GJ (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 NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of molecular interaction networks. Genome Res 13: 2498–2504

© 2010 EMBO and Macmillan Publishers Limited
