


Ghent University – Faculty of Sciences

Department of Plant Biotechnology and Bioinformatics

VIB – Department of Plant Systems Biology

CropDesign N.V.

Building a functional interactomics approach to enhance growth or seed yield in rice

Maarten Dedecker

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor in Sciences: Biotechnology

September 2015

Promoters: Prof. Dr. Geert De Jaeger & Dr. Steven Vandenabeele

Dit onderzoek werd uitgevoerd gedurende de periode 2010-2015, deels in het departement Planten Systeembiologie van het Vlaams Instituut voor Biotechnologie (VIB) en de Universiteit Gent, deels in CropDesign N.V.

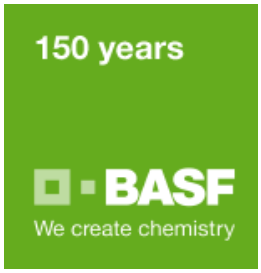
This work was conducted during the period 2010-2015, partly in the department of Plant Systems Biology of the Flanders Institute for Biotechnology (VIB) and the Ghent University, and partly in CropDesign N.V.

Dit onderzoek werd gefinancierd door het agentschap voor Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen, via een Baekelandt mandaat: IWT100222).

This work was supported by the Agency for Innovation by Science and Technology in Flanders (IWT-Flanders, through a Baekeland grant: IWT100222).

agentschap voor Innovatie
door Wetenschap en Technologie

cropdesign



Board of examiners

Promoters

Prof. Dr. Geert De Jaeger	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University
Dr. Steven Vandenabeele	CropDesign N.V.

Examination committee

Prof. Dr. Ann Depicker (Chair)	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University
Prof. Dr. Bart Devreese	Department of Biochemistry and Microbiology Faculty of Sciences Ghent University
Prof. Dr. Kris Gevaert	VIB Department of Medical Protein Research Department of Biochemistry Faculty of Medicine and Health Sciences Ghent University
Dr. Valerie Frankard	CropDesign N.V.
Dr. Hilde Nelissen	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University

Dr. Laurens Pauwels	VIB Department of Plant Systems Biology Department of Plant Biotechnology and Bioinformatics Faculty of Sciences Ghent University
Dr. Srijeet Mitra	BASF Plant Science, RTP

Table of content

List of abbreviations

5PTase13	myo-inositol polyphosphate 5-phosphatase
ABA	abscissic acid
ABI3	ABSCISSIC ACID INSENSITIVE 3
ACE	ACTIVATOR FOR CELL ELONGATION
ADP	adenosine diphosphate
AGO10	ARGONAUTE 10
AGPase	adenosine diphosphoglucose pyrophosphorylase
AIF	ATBS1-INTERACTING FACTOR
AMPK	AMP-ACTIVATED PROTEIN KINASE
AN3	ANGUSTIFOLIA 3
APC	anaphase promoting complex
APC10	ANAPHASE PROMOTING COMPLEX SUBUNIT 10
APG	ANTAGONIST OF PGL1
AP-MS	affinity purification coupled to mass spectrometry
ARGOS	*AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE*
AS tag	TAP tag, consisting of CBP, TEV cleavage site and tandem repeat of the ZZ domain of protein A
ASC domain	association to the complex domain
ATBS1	ACTIVATION-TAGGED SUPPRESSOR OF BRI1-103
BAC	bacterial artificial chromosome
BEE	BR ENHANCED EXPRESSION
bHLH	basic helix-loop-helix
BIM	BES1-INTERACTING MYC LIKE
BNQ	BANQUO
BPE	BIG PETAL
BR	brassinosteroid
brd1	*brassinosteroid dependent 1*
BRI1	*BRASSINOSTEROID INSENSITIVE 1*
BZR1	BRASSINAZOLE-RESISTANT 1
Bt crops	crops expressing the insecticidal *Bacillus thuringiensis* δ-endotoxins
BU1	BRASSINOSTEROID UPREGULATED 1
bZIP	basic leucine-ZIPPER
CaMV 35S	35S promoter of the cauliflower mosaic virus
CBM	carbohydrate-binding module
CBP	calmodulin binding peptide
CBS	cystathionine β-synthase
CcdB	control of cell death b
CCS52A	CELL CYCLE SWITCH PROTEIN 52A
CDC20	CELL DIVISION CYCLE 20
CDK	cyclin-dependent kinase
CDKD	CYCLIN-DEPENDENT KINASE D
Cgreen tag	TAP tag, consisting of CBP, TEV cleavage site and GFP
ChAP	chromatin affinity purification
ChIP	chromatin immunoprecipitation
CIB1	CRYPTOCHROME INTERACTING BHLH 1
CIPK1	CALCINEURIN B-LIKE-INTERACTING PROTEIN KINASE 1
CKI	CYCLIN-DEPENDENT KINASE INHIBITOR
CKS1	CYCLIN-DEPENDENT KINASE REGULATORY SUBUNIT 1
Cm^R	chloramphenicol resistance
COP1	CONSTITUTIVE PHOTOMORPHOGENIC 1
CRISPR	clustered regularly interspaced short palindromic repeats
CRL	CULLIN-RING ubiquitin E3 ligases

CRL4A	CULLIN4A-RING
CS tag	TAP tag, consisting of CBP and SBP
CWIN	cell wall-located invertase
CYC	cyclin
DCAF	DDB1-CUL4 ASSOCIATED FACTOR
DDB1A	UV-DAMAGED DNA BINDING PROTEIN A
DNA	deoxyribonucleic acid
EGTA	ethylene glycol tetraacetic acid
EXP	EXPANSIN
FLC	FLOWERING LOCUS C
FRET/FLIM	fluorescence resonance energy transfer/fluorescence lifetime imaging
G6P	glucose-6-phosphate
GA	GIBERELLIC ACID
GCN5	GENERAL CONTROL OF AMINO ACID SYNTHESIS 5
GFP	green fluorescent protein
GIF1	GRF-INTERACTING FACTOR 1
GLK	GOLDEN2-LIKE
GLN1-3	*GLUTAMINE SYNTHETASE 1-3*
GM	genetically modified
GNAT	GCN5-related N-acetyltransferases-like
GO	gene onthology
GRF5	GROWTH REGULATING FACTOR 5
GS tag	TAP tag, consisting of SBP, TEV cleavage site and tandem repeat of the ZZ domain of protein G
GSgreen tag	TAP tag, consisting of SBP, TEV cleavage site and GFP
GS^{rhino}	TAP tag, consisting of SBP, HR3C cleavage site and tandem repeat of the ZZ domain of protein G
GW6	grain weight 6
GWAM	genome wide association mapping
H2B	HISTONE 2B
H_2O_2	hydrogen peroxide
HA	hemagglutinin
HBI1	HOMOLOG OF BEE2 INTERACTING WITH IBH1
HBT	HOBBIT
HFR1	LONG HYPOCOTYL IN FAR RED 1
HMG-CoA reductase	3-hydroxy-3-methylglutaryl-coenzyme A reductase
HR3C	human rhinovirus 3c
HY5	light-responsive element
IBH1	INCREASED LEAF INCLINATION1 BINDING BHLH
IgG	immunoglobulin G
ILI	INCREASED LEAF INCLINATION
INV	invertase
IPT	isopentenyl transferase
iTRAQ	isobaric tags for relative and absolute quantitation
KA1	kinase assocated 1 domain
kb	kilobases
KDR	KIDARI
KIS	kinase-interacting sequence
Km^R	Kanamycin resistance
LAP tag	localisation and affinity purification tag
LB	left border of the T-DNA
LC	liquid chromatography
LRE	Light-Responsive Element
MAIF1	MIRNA'S REGULATED AND ABIOTIC STRESS INDUCED F-BOX 1

MAT1	MENAGE A TROIS
Mmi	MEIOTIC MRNA INTERCEPTION
MS	mass spectrometry
MS1	mass spec survey scan
MS2	mass spec fragment spectrum
MSU	Michigan State University annotation database
NADPH	nicotinamide adenine dinucleotide phosphate
NADP-ME	NADP-Malic enzyme
NADP-MPDH	NADP malate dehydrogenase
NAP	NUCLEOSOME ASSEMBLY PROTEIN
NSAFs	normalized spectral abundance factors
OA	oxaloacetate
ORF	open reading frame
OsBRI1	*Oryza sativa* BRASSINOSTEROID INSENSTIVE 1
OsD1/RGA1	D1/ rice heterotrimeric G-protein α subunit
OsFIE2	*Oryza sativa* FERTILIZATION-INDEPENDENT ENDOSPERM 2
OsIHAT1	*Oryza sativa* INTRINSIC HISTONE ACETYL TRANSFERASE ACTIVITY 1
OsO3L	*Oryza sativa* OXIDATIVE STRESS 3-LIKE
OsPTR9	*Oryza sativa PEPTIDE TRANSPORTER 9*
OXS	OXIDATIVE STRESS 3
PAR	PHYTOCHROME RAPIDLY REGULATED
PCA	protein complementation analysis
PEL	PSEUDO-ETIOLATION IN LIGHT
PEP	phosphoenol pyruvate
PEP	plastid-encoded RNA polymerase
PEPC	phosphoenol pyruvate carboxylase
PGL	PROMOTER OF GRAIN LENGTH
PIF	PHYTOCHROME INTERACTING FACTOR
PP2C	protein 2C phosphatase
PPDK	pyruvate phosphate dikinase
PPI	protein-protein interaction
PRE	PACLUBUTRAZOL RESISTANCE
PRL1	PLEITROPIC REGULATOR LOCUS 1
PRO_{35S}	35S promoter of the Cauliflower mosaic virus
PRO_{GOS2}	promoter of the rice GOS2 gene
PRO_{HMG}	HIGH MOBILITY GROUP PROTEIN promoter
QTL	quantitative trait locus
RAP	rice annotation project annotation database
RB	right border of the T-DNA
RING	REALLY INTERESTING NEW GENE
RNA	ribonucleic acid
RubisCo	ribulose bisphosphate carboxylase
RuBP	ribulose-1,5-bisphosphate
SAG12	SENESCENCE-ASSOCIATED GENES 12
SAUR	SMALL AUXIN UPREGULATED
SBP	streptavidin-binding peptide
SCC4	single cell C4 species
SCF	SKP1-CULLIN-F-BOX
SEC-MS	size-exclusion chromatography coupled to mass spectrometry
SFZZ tag	TAP tag, consisting of StrepIII, FLAG, TEV cleavage site and tandem repeat of the ZZ domain of protein A
sg1	*short grain 1*
SH tag	TAP tag, consisting of SBP, TEV cleavage site and HA
SIG	sigma factor genes
SIIIgreen	TAP tag, consisting of SIII, TEV cleavage site and GFP

SILAC	stabile isotope labelling by amino acids in cell culture
SKIN	SNRK1-INTERACTING NEGATIVE REGULATORS
SKP1	S-PHASE KINASE ASSOCIATED PROTEIN 1
SnAKs	SnRK1 activating kinases
SNF1	SUCROSE NON-FERMENTING
SNP	single nucleotide polymorphism
SnRK1	SUCROSE NON-FERMENTING-RELATED KINASE 1
SPL	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE
SuS	sucrose synthase
T6P	trehalose-6-phosphate
TALE	transcription activator-like effector
TAP	tandem affinity purification
TAPa tag	alternative TAP tag
TAPi tag	improved TAP tag
TBP	TATA-BOX BINDING PROTEIN
tChAP	tandem chromatin affinity purification
TChAP	tandem chromatin affinity purification
TCTP	TRANSLATIONALLY-CONTROLLED TUMOR PROTEIN
T-DNA	transfer-DNA
TEV	tobacco etch virus
TFIIH	general transcription factor IIH complex
TIC	total ion current
TKW	thousand kernel weight
TMO7	TARGET OF MONOPTEROS 7
TPS	trehalose-6-phosphate synthase
TSI1	TOBACCO STRESS-INDUCED 1
TSIP1	TSI1-INTERACTING PROTEIN
UBA	ubiquitin-associated domain
UDP-glucose	uridine-diphosphoglucose
UPB1	UPBEAT 1
VRN	VERNALIZATION
XPB	XERODERMA PIGMENTOSUM B
XPD	XERODERMA PIGMENTOSUM D
Y2H	yeast two-hybrid
YFP	yellow fluorescent protein

Part I. Scope

Chapter 1: Scope

CropDesign holds a unique position on world scale. Owing to its unique TraitMill phenotyping platform, the company possesses an enormous database linking genes to specific phenotypes related to growth or seed yield in rice. These genes are involved in various processes, and their inter-relationship is mainly unknown. In most cases, the actual cause of how these genes enhance growth or seed yield is even unclear. To unravel the mechanisms that define growth or seed yield at the molecular level, it is necessary to understand how the molecular machinery that determines yield parameters operates.

One important data layer that provides clarity in a cell's working mechanisms is the unravelling of protein-protein interactions that support genes in their function. Indeed, proteins rarely act as single entities. Instead they co-operate, mainly through protein-protein interactions, to function as molecular machines.

The main objective of this project was therefore to build a TAP technology platform in rice. Rice is the preferred organism to work with as 1) CropDesign's yield 'lead' genes were discovered in rice, 2) rice as a crop is of great commercial importance on world scale and 3) rice can be used as a model crop for corn, wheat and other important cereal crops.

A first objective is **to develop the optimal tools to perfection TAP in rice**. For this, we will test different TAP tags and promotors for their efficiency in rice callus by using the rice ortholog of CYCLIN-DEPENDENT KINASE REGULATORY SUBUNIT 1 (CKS1) as bait protein. CKS1 has a conserved role in cell cycle regulation as scaffold protein. We will also produce the necessary Gateway® vectors to obtain high cloning flexibility. Our goal is to find a TAP tag/promotor combination that allows obtaining higher yield and specificity in purifying complexes than the earlier published TAPi tag.

A second goal is **the establishment of a functional TAP platform in rice callus**. The tools to perfection TAP in rice will be tested in rice callus. For this, rice CKS1 will be used as a TAP 'bait' protein. The interactors of CKS1 are known for Arabidopsis and the gene is strongly conserved between Arabidopsis and rice.

The combination of an optimal TAP tag/promotor with the use of the ultrasensitive Orbitrap mass spectrometry offers the possibility to efficiently purify and characterize protein complexes from rice plant tissues. We therefore aimed to **develop a functional TAP platform in rice plants**.

Once we'll have the tools for an optimised TAP protocol in callus and plant tissues at hand, the final target is to **identify protein complexes from multiple lead baits using the rice callus platform and the rice plant TAP platform**. The current state of the art of TAP in rice detects interactors for only 23% of the screened baits. Our TAP platform should result in a higher throughput and a higher success rate compared to this state of the art.

Part II. Introductory chapters

Chapter 2: Improving yield: how to pimp my crop?

Modern agriculture requires yield increases

Food or in general nutrition is one of the basic needs for any organism. The advent of agriculture, the cultivating of domestic plants, allowed fewer people to provide more food. The stability that came with regular, predictable food production led to a tremendous increase in population density. This enabled people to do more than hunt for daily food requirements: they started to travel, trade, and communicate. Continuous improvements in agricultural technologies and practices allowed sustaining this population growth. E.g. the introduction of the iron plough, crop rotation, synthetic fertilizers and more recently the first Green Revolution provided the necessary quantum leaps in agricultural yields. Nowadays, demographic projections predict the world population to reach about 9.6 billion people by 2050. This population increase and the prospected changes in life style – more people tend to eat meat – imply that global agricultural productivity should increase by 60% to sustain proper human nutrition without significantly increasing the cultivated area [1]. The latter implies that with the natural degradation of agricultural land, simply augmenting planting of crops is no longer a viable solution. Most of the increase in production will thus have to be achieved by increasing the yield per area [1]. This poses an enormous pressure on the agricultural industry to produce more output than ever before.

Today's ways to enhance yield from crops

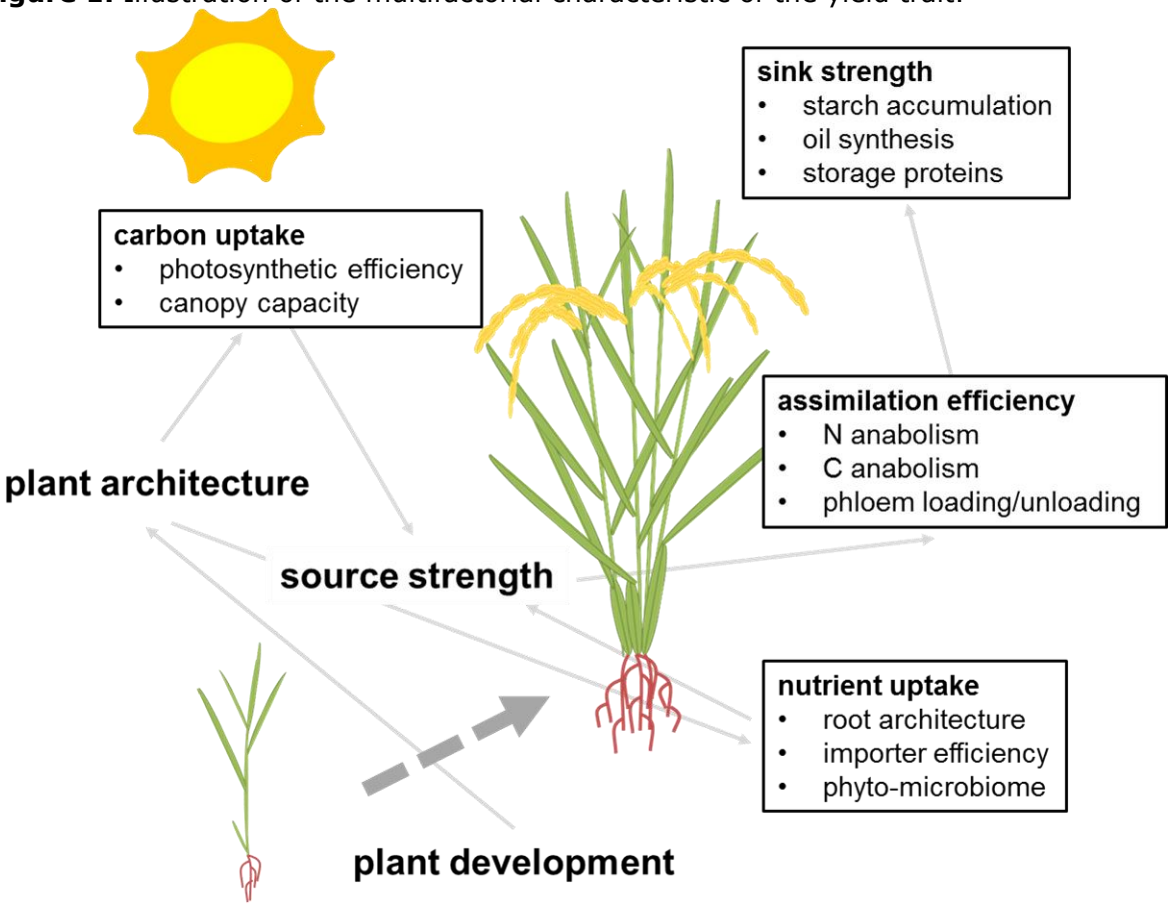
A way to enhance yield is to improve the varieties themselves for higher resilience and/or higher yield potential. This can be achieved by breeding, exploitation of heterosis (through the use of hybrids) and/or genetic modification. The latter offers the huge advantage that it adds an enormous potential of novel (transgenic) variation to supplement the natural variation for yield or stress resistance traits. To date, commercially available genetic engineered crops are mainly focused on resilience against biotic stress. Herbicide tolerant GM cultivars (e.g. Clearfield® and Roundup Ready® crops) are less susceptible to weed competition and have been very successful in maize, soybean and canola. Transgenic insect resistance (e.g. crops expressing the insecticidal *Bacillus thuringiensis* δ-endotoxins, or 'Bt crops') showed to be very effective in maize and cotton. Also transgenic resistance to fungal diseases is being developed (www.croplife.org). In 2013, the first transgenic crop that targets yield maintenance under abiotic stress - in this case drought stress - was introduced into the market by the Monsanto/BASF collaboration (Genuity® DroughtGard™ Hybrids Corn, [2]).

These higher-resilience traits are extremely suitable in closing yield gaps caused by the (a)biotic stresses they target. However, once established, the yield increases delivered by controlling pests, disease and weeds cannot be repeated. At best they can be maintained [3]. Therefore, continued improvements in crop genetics are required to ensure potential yields into the future. Also here, genetic modification can greatly aid in adding genetic diversity by breaking the genetic boundary of both genes and their spatial and temporal control of expression by promoters.

Multiple routes for boosting intrinsic yield

To further boost plant yield, it will be necessary to understand how the molecular machinery that determines yield parameters operates. Due to the multifactorial nature of the complex trait 'yield', many aspects fall in this scope (Figure 1). In first instance, maximal exploitation of the plant's energy sources can be targeted. Nutrients taken up through the root system and carbohydrates produced during photosynthesis in the leaves are mobilized throughout the plant to sustain its growth. Both nutrient uptake and photosynthesis efficiency are therefore excellent targets for enhancing plant yield. Not only the capturing, but also the distribution of assimilates to the tissues that require these is a potential target to improve plant yield. This can be achieved through optimisation of assimilate transport and enhancement of sink capacity. Alternatively the orchestration of plant growth could be targeted to aim for enhancing growth or seed yield, and/or optimizing plant architecture for high density cultivation.

Figure 1. Illustration of the multifactorial characteristic of the yield trait.



The main developmental and physiological processes targeted for yield enhancement described in this chapter are shown. The complexity of the trait is illustrated by the interdependency of the different parameters underlying yield, as illustrated by the grey arrows. N: nitrogen; C: carbon.

Plants rely for their nutrient supply mainly on what mineral nutrients are available in the soil. Nitrogen, phosphorus, magnesium and potassium are the building blocks of crucial cellular components, like proteins and nucleic acids, and are therefore required in significant quantities. Other minerals, like iron, zinc,

manganese and copper are usually chelated or used as cofactors for enzyme activity and therefore only needed in small amounts. As non-motile organisms, plants are forced to utilize a plethora of sophisticated mechanisms to acquire sufficient amounts of these nutrients for proper growth. These mechanisms include optimizing root structure to better mine the soil, induction of high affinity transporters and the establishment of symbioses for facilitated nutrient uptake. For agriculture purposes, engineering of more efficient nutrient uptake can lead to plants that make better use of the available fertilizers, leading to higher crop yields. This in principle only holds true if the use of sufficient fertilizer is guaranteed. In that case, transporters are an excellent route to increase nutrient uptake from the soil [4]. Indeed, altering the expression of *OsPTR9* (for *PEPTIDE TRANSPORTER 9*), a rice nitrate transporter, was shown to significantly affect nitrogen uptake and grain yield under limiting and normal N fertilization in the field [5]. Increased nutrient assimilation efficiency can also be achieved by optimizing the intracellular assimilation and recycling process. A known target related to this strategy is glutamate synthetase, an enzyme that catalyzes the condensation of glutamate and ammonia to form glutamine. Overexpression of *GLN1-3* (for *GLUTAMINE SYNTHASE 1-3*) in leaves led to 30% increase in kernel number in maize [6]. Also in wheat and Arabidopsis, an increase in yield or biomass was obtained by overexpressing this enzyme [7–9]. Similarly, rice plants overexpressing alanine aminotransferase, catalysing the reversible transfer of an amino group from glutamate to pyruvate to form 2-oxoglutarate and alanine, had significantly increased biomass and grain yield [10].

Plants use the energy of the sun to produce biomass during photosynthesis. They absorb specific wavelengths (400-700 nm) of sunlight using the pigment chlorophyll, to convert sunlight to chemical energy. This energy is used to 'fix' or convert atmospheric CO_2 into organic compounds. Experiments increasing photosynthesis by artificial elevation of [CO_2] provide evidence that increasing photosynthesis in a crop (in this case soy) under standard field conditions does result in an increase in yield of 15% [11]. To achieve similar yield increase without having to increase [CO_2], different possible targets could be selected. First of all, the enzyme taking care of the first committed step in carbon fixation, ribulose bisphosphate carboxylase (RubisCo), evolved from an environment that initially contained no oxygen. As a consequence, RubisCo has low selectivity in distinguishing CO_2 from O_2 in the carboxylation step of ribulose 1,5-diphosphate. Especially in C3 plants, photosynthetic efficiency is dramatically reduced because of oxygen being combined with the sugar instead of CO_2 in a process called photorespiration. In addition, the selectivity of RubisCo is temperature sensitive and becomes poorer at higher temperatures. Net losses of carbon due to this reduced photosynthetic efficiency are estimated at 25% [12]. C4 plants use a compartmentalization system to concentrate the CO_2 levels around RubisCo, favouring the carboxylation reaction and allowing higher photosynthetic efficiency (and higher yields) at high temperatures [13]. In fact, this also results in a net increase in water and nitrogen use efficiency [13,14]. The storage of CO_2 in the bundle-sheath cells results in a lowering of the CO_2-levels in the mesophyll cells, contributing to a swifter diffusion of CO_2 in the leaf. Stomata can thus be smaller and/or less numerous, preventing water loss due to transpiration. Since RubisCo is more efficiently used in C4 plants, less of the protein has to be produced, saving

nitrogen resources [13,14]. There have been many attempts undertaken to introduce (parts of) this more efficient C4-like photosynthetic system in C3 crops, such as rice (http://c4rice.irri.org/) and potato [15].

Another route for increasing photosynthetic efficiency is adapting the plant's development as such that the time window of canopy closure (and thus photosynthetic capacity) is maximized. A comparison of maize with a relative C4 grass, the cold-tolerant perennial *Miscantus x giganteus,* showed that the latter produced 60% more biomass [16]. This higher productivity is accounted to a 4 weeks earlier closing of the canopy and a 4 weeks later senescence. Especially delaying senescence is thoroughly studied [17]. For example a chimeric construct of the isopentenyl transferase (IPT) encoding gene controlled by the senescence promoter PRO_{SAG12} was developed, allowing senescence-specific production of cytokinin [18]. The resulting senescence inhibition showed a positive effect on photosynthesis and productivity of the entire canopy in various crop species. While the long vegetative period showed to be beneficial for biomass production, it might however be detrimental for grain yield. Over-prolonged leaf longevity may prevent effective nutrient recycling and a good balance between delayed leaf senescence and nutrient translocation to sink tissues is probably needed for increasing yield in crops [19].

Analogous to what was discussed related to nutrient supply, carbon assimilation efficiency can also be achieved by optimizing the intracellular assimilation by heterotrophic sink tissues. Indeed, carbon fixated in the photosynthetic tissues is translocated in the form of sucrose to support non-photosynthetic tissues. Sink capacity can be strengthened by either altering the biochemistry of the sink organs or enhancing the sucrose phloem loading/unloading capacity. Once the sucrose reaches the sink cells, it is either hydrolysed by invertase (INV) to glucose and fructose, or degraded by sucrose synthase (SuS) into uridine-diphosphoglucose (UDP-glucose) and fructose. Overexpression of a cell wall-located invertase (*CWIN*) in rice grains driven by its native promoter led to increased grain size [20]. Similarly, a constitutively expressed *CWIN* improved grain yield up to 145.3% in transgenic maize plants as compared to the wild-type plants [21]. Overexpression of SuS in cotton reduces seed abortion, resulting in more mature seeds and higher cotton fibre yield [22]. More downstream of the SuS pathway, overexpression of the adenosine diphosphoglucose pyrophosphorylase (AGPase), which catalyses the first committed step into starch biosynthesis, was targeted. AGPase is a heterotetramer defined by 2 small and 2 large subunits. Overexpression under the control of an endosperm-specific promoter of either both genes encoding for the 2 small subunits or the genes for the 2 large subunits enhanced seed weight and starch content in maize. The 100-grain weight increased 15% over the wild type [23]. Also trehalose-6-phosphate (T6P) plays a crucial role, not only in sugar metabolism, but also in signalling and plant development. Its importance is reflected in a central role in carbon metabolism and plant development. Biosynthesis of the compound consists of a condensation reaction consuming UDP-glucose and glucose-6-phosphate (G6P) catalysed by trehalose-6-phosphate synthase (TPS). UDP-glucose is a known cell wall, cellulose and callose precursor, whereas G6P is the starting point for energy metabolism generating NADPH (for nicotinamide adenine dinucleotide phosphate) for reductive biosynthesis. Recent

studies show that TPS can sense sucrose availability to generate T6P as a signal to promote growth and starch and cell wall biosynthesis. This could be achieved through the inhibitory effect of T6P on the activity of SUCROSE NON-FERMENTING-RELATED KINASE 1 (SnRK1) [24,25]. Under restrictive growth conditions but with abundant sucrose, T6P makes plants ready to recover by stimulating expression of growth-related genes [26]. T6P also plays pivotal signalling roles in regulating stress responses [25] and development, including embryogenesis and flowering. T6P is therefore a sensitive target for genetic engineering to improve crop yield, which will require highly targeted approaches to manipulate T6P metabolism very precisely in time, space and magnitude within plant tissues.

The importance of plant architecture in agriculture is reflected in the breeding efforts during the green revolution. Semi-dwarf rice and wheat varieties were introduced at that time. This prevents diversion of resources towards stem elongation in field conditions (high density growing with high nitrogen input) that normally favour this process, and avoids lodging. According to the International Rice Research Institute, the ideal rice plant architecture consists of low tiller (branches carrying inflorescence) numbers, a high number of productive tillers, 200-250 grains per panicle (the cluster of flowers at the top of the plant), dark-green thick and erect leaves and a vigorous and deep root system. The recent identification of a discrete number of genes as major players in the increase in plant productivity during the Green Revolution can help in further developing rice towards this ideal architecture [7]. In this frame IDEAL PLANT ARCHITECTURE and WEALTHY FARMER'S PANICLE QTL's were identified, correlating with much of the variation in plant architecture in two rice varieties (respectively Shaoniejing and ST-12 varieties) [27,28]. In both cases, fine mapping of the QTL led to the identification of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 14 as the gene that can affect plant architecture. Transgenic lines overexpressing OsSPL14 have fewer tillers with stronger culms (stalks) as well as an increased number of panicle branches and increased grain yield [27,28]. Reduced levels of OsSPL14 to the contrary, lead to shorter plants with more tillers and less grains [27].

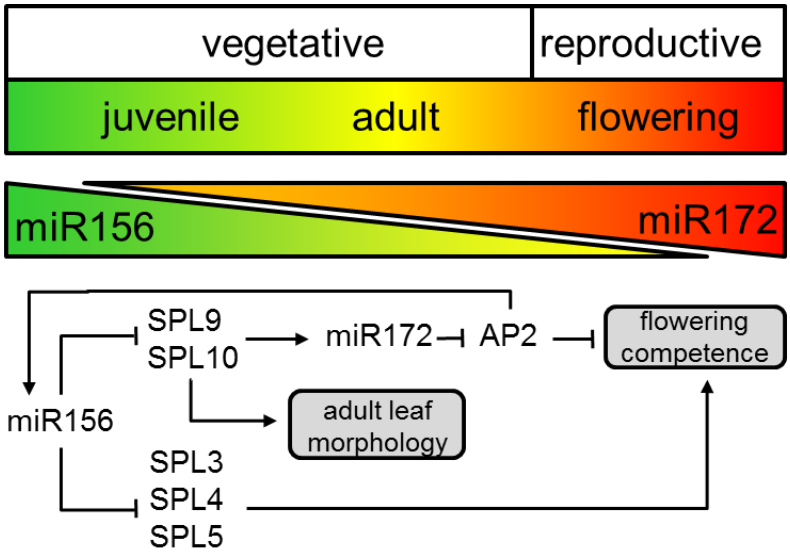
Unlike rice and wheat, corn did not went through a similar yield transformation as the Green Revolution. The development of dwarfed corn varieties with unchanged kernel yield could therefore open new perspectives [29]. The simple change to dwarfing in oilseed rape (*Brassica napus*) through genetic engineering illustrates that yield increase through targeted manipulation is within reach [30].

A comprehensive approach for enhancing yield potential is altering the plant's development to achieve enhanced growth or seed yield. For this, Arabidopsis and rice are the most common model systems to find genes and regulatory networks controlling yield determining processes. In turn, the knowledge from these models can be used for further growth improvement in other species including cereals. The formation of organs or tissues is studied as a proxy for whole plant development. In Arabidopsis research, focus is mainly on unravelling the mechanisms behind organ and more particular leaf growth. The general believe is that organ growth at the cellular level is determined by spatiotemporal-specific cell proliferation and cell differentiation. In a first phase cell proliferation takes place, followed by cell elongation. These two events occur at different times during

development and in different cell types, each with a different rate. The co-ordination of both processes is thus both crucial and extremely complex. A variety of organ growth regulators were identified streamlining either or both processes at different stages throughout leaf development. Some of these genes resulted in positive effects in the field. For example the *Zea mays* ortholog of *ARGOS* or *AUXIN REGULATED GENE INVOLVED IN ORGAN SIZE*, was shown to enhance maize organ growth, grain yield and drought stress tolerance [31]. Also, key regulators controlling seed yield were identified. Seed yield is clearly a lot more complicated trait than organ size. Not only seed size, but also the number of seeds, defined by inflorescence architecture and floral development, and seed filling play pivotal roles in the final yield (calculated as total weight of seeds per plant). In addition, seed number is often negatively correlated with seed size [32]. A recent effort in Arabidopsis assembled the fragmented knowledge on seed yield into a global and coherent picture on 46 genes [33]. The study revealed extensive variation for total rosette area and the five seed yield parameters studied, complicating the interpretation of the biological processes behind the trait. First of all, not all results showed to be reproducible, even by using plants with the same genotype. This stresses that different (lab) environmental conditions can influence experimental reproducibility. Although leaf size was found not correlated with seed yield, some of the genes positively regulating seed yield were found to also control leaf growth [33]. Indeed, *GRF1*, *GRF5* (for *GROWTH-REGULATING FACTOR 1* and *5* respectively) and *BRI1* (for *BRASSINOSTEROID INSENSITIVE 1*) overexpressors, all affecting leaf size, were found to produce larger seeds. Overexpression of the GRF transcription factors produced larger cotyledons containing larger (*GRF1*) and more (*GRF5*) cells [34]. The increased seed size observed in these plants possibly resulted from the production of larger embryos containing larger cotyledons. The role of brassinosteroids in seed development is stressed by *BRI1* overexpression affecting seed size. This is not surprising since previous links to brassinosteroid signalling to seed size were reported (reviewed in [35]).

Advances in understanding the transition phases during plant development could also provide pivotal information to increase yield [36]. This could for example enable to engineer plants that have short vegetative phases and longer reproductive phases, leading to higher seed yields. A potential target for this is the antagonistic action of miR172 and miR156 miRNAs and their respective targets on plant development [36] (Figure 2). Briefly, during early development the levels of miR156 are initially high, promoting the juvenile vegetative growth phase in Arabidopsis seedlings [37]. As the plant matures, the levels of miR156 steadily decrease, allowing for the production of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 (SPL9) and SPL10 proteins that promote adult leaf traits [37]. Simultaneously, SPL9 and SPL10 directly induce the expression of miR172 genes [38]. Increased levels of miR172 result in the down regulation of six APETALA2-like transcription factors that normally repress flowering. Release from this repression, in combination with the flower-promoting actions of SPL3, SPL4 and SPL5, makes the plant competent to flower and enables transition to flowering [38].

Figure 2. Regulation of transition phases in Arabidopsis.



During early development the levels of miR156 are initially high, promoting the juvenile vegetative growth phase in seedlings. As the plant matures, miR156 levels steadily decrease, allowing for the production of SPL9 and SPL10 proteins that promote adult leaf traits. At the same time, SPL9 and SPL10 directly induce the expression of *MIR172* genes. Increased levels of miR172 result in the downregulation of six AP2-like transcription factors (denoted as 'AP2') that normally repress flowering. Release from this repression, in combination with the flower-promoting actions of SPL3, SPL4 and SPL5, makes the plant competent to flower and the transition to flowering can occur. Abbreviations: AP2, APETALA2; miR156, mature miRNA156; miR172, mature miRNA172.

Genes are also tested in high-throughput fashion to verify if their altered expression effects plant yield. This can provide additional information on genes and pathways involved in plant development and yield biology. Most high throughput testing is performed by private companies such as Mendel Biotechnology, which tests mainly transcription factors in Arabidopsis (www.mendel.com), and CropDesign. The latter developed TraitMill, a high-throughput platform that enables large-scale transgenesis and plant evaluation. TraitMill allows testing the effect of genes and gene combinations on plant phenotypes. It can be used to successfully evaluate hundreds of independent promoter-gene combinations per year, either under optimal growth conditions or under different abiotic or nutrient stress regimes. The TraitMill platform operates in rice and is specially designed to measure alterations in growth with high sensitivity [39]. Although slower and more costly as a testing system than Arabidopsis, rice has the advantage that it is a crop itself. It can serve as a more suitable model for cereal crops, being evolutionary more close than Arabidopsis. The potential of the TraitMill platform is determined by the quality of the genes that are tested. With the abundant availability of genomic sequences for plants and the throughput of approximately 1000 constructs a year, it is a continuous challenge to look for innovative ways of making the best gene choices for the TraitMill platform. CropDesign possesses a unique databank with more than 500 lead genes that significantly contribute to an enhancement in at least one of the growth and/or yield components under normal or stress conditions. This extensive

databank enables CropDesign to fulfil a pioneering role in the unravelling of molecular networks involved in yield.

Identifying the connections between growth regulators might unravel the complexity of yield

Despite these numerous genes affecting yield identified, our understanding of yield biology is still scarce compared to its overwhelming complexity. An illustration of our current abilities is nicely provided in the engineering of *Brassica napus* flowers [40 and references within]. These turn bright yellow from the petals held above the leaves. The problem is that these petals can block nearly 60% of the photosynthetic effective light and facilitate fungal pathogen entry. From the well-established flower organ specification models from Arabidopsis research, canola lines were generated with petals converted into sepals, which are smaller and block only little light. This led to increased photosynthesis and higher yield. Subsequently, a line was found completely lacking petals with no other deleterious effects. This line was mutated in a previously unstudied gene, which function is still not understood. That a mutation in an unknown gene would confer such a desirable phenotype in the best-studied developmental genetic pathway exposes the hiatuses in our understanding of how genes generate phenotypes.

Apart from knowing all individual growth regulating components, it is clear that only a tight co-ordination between these numerous players can support the formation of a functional organ. This is illustrated by compensatory responses among yield components that sometimes prevent to reach an overall positive yield effect [41]. To face this challenge, a better understanding is required of the molecular processes that drive intrinsic yield [34], together with their impact on the organ level and plant level [40]. The connections between the components of the network and their dynamics can be further disentangled through their identification. The general trend within the community is therefore to also test 'stacked' constructs, combinations of independent alleles with the aim to achieve additional or synergistic effects. A number of efforts on assaying epistatic interactions between independent alleles were done in Arabidopsis [42–44]. In the most exhaustive study, a pairwise combinatorial screen between 13 transgenic Arabidopsis lines with an increased leaf size was performed. The finding was that a large proportion of the combinations analysed showed an additional increase in leaf size resulting from a positive epistasis on growth [42]. This study clearly shows the potential of combinatorial approaches to discover further the connections of growth regulatory networks. Studies in yeast have shown that most genetic interactions occur between genes involved in the same biological process, except for highly connected genes [45,46]. In agreement with this finding, the study in Arabidopsis found that combining a *GIF1* (for *GRF-INTERACTING FACTOR 1)* overexpressor with a *GRF5* overexpressor increased leaf size more than expected [42]. GIF1 and GRF5 were indeed found to interact in a yeast two-hybrid assay [47]. Similarly, another study showed AtbZIP10 and AtbZIP25 to interact *in vivo* with ABI3 (for ABSCISSIC ACID INSENSITIVE 3), an important regulator of gene expression in the seed of Arabidopsis. While transgenic plants ectopically overexpressing *AtbZIP10*, *AtbZIP25*, and *ABI3* revealed that none of these factors could significantly activate a reporter gene individually, co-overexpression of

AtbZIP10/25 with *ABI3* resulted in a remarkable increase in the activation capacity over the reporter, suggesting that they are part of a regulatory complex involved in seed-specific expression [48]. Similarly, transgenic plants co-expressing *TOBACCO STRESS-INDUCED 1* (*TSI1*), a transcription factor that plays an important role in both biotic and abiotic stress signalling pathways, and its interactor *TSI1-INTERACTING PROTEIN 1* (*TSIP1*) showed to display stronger pathogen resistance and salt tolerance than did transgenic plants expressing either *TSI1* or *TSIP1* alone. Concurrent with this, the expression of a subset of stress-related genes was induced in a cooperative manner in *TSI1/TSIP1* transgenic plants [49].

Protein-protein connections *as such* also provide a great deal of knowledge on the resulting phenotype. Soon after publication of the first proteome-scale human interactome network map [50,51], it was observed that proteins involved in the same disease are more connected to each other than expected by chance and often form densely connected modules [52]. Based on this finding, network approaches were successfully developed to understand the connectivity of known disease proteins and identify new ones. Interactome analysis focused on proteins implicated in ataxia for example revealed not only that ataxia-causing genes interact more than expected by chance [53], but also that proteins encoded by genetic modifiers of disease risk were found as interaction partners of the disease proteins [54]. Similar interactome analyses were successful in identifying novel causative genes for a number of human diseases, including schizophrenia [53,55], Huntington's disease [56] and breast cancer [57]. This finding can be extrapolated to plants and the phenotype 'yield'. The identification of interaction partners of yield enhancing gene products, might lead to the discovery of new yield enhancing genes and give more insight on the connections within a 'yield enhancing network'. The recent effort to unveil the cell cycle machinery for example [58–60], provided a myriad of new potential growth regulators in Arabidopsis [61].

Protein interaction data was also employed to analyse results from genome wide association mapping (GWAM) studies. These studies identify genomic loci, which contain a variable amount of genes that show association with complex phenotypes as disease susceptibility in humans or crop yield in plants. It is however often difficult to exactly pinpoint correlated genes within a given locus. In addition, loci without genome-wide statistical significance may still contain variants affecting the phenotype. Including information from protein interactions is a potential powerful way to address these challenges [62–65]. For example, the combination of experimental and computational network analysis of Alzheimer's disease susceptibility proteins demonstrated that Alzheimer susceptibility proteins exhibit increased interconnectivity allowing identification of several novel Alzheimer candidate genes in GWAM loci [66]. In a similar interactome analysis of multiple sclerosis GWAM data, a possible role for axon guidance and synaptic potentiation was revealed [62]. These studies illustrate that interaction data hold great potential for helping to interpret genetic data for complex phenotypes. Basically, GWAM studies examine single nucleotide polymorphisms (SNPs) across the genome. A part of these SNP's that are correlated to the complex phenotype might relate to changes in amino acids. For a long time, it is known that disease causing point mutations can act through disruption of physical interactions [67]. In a recent

large-scale analysis of disease mutations interactome information was integrated with data on disease-causing mutations and structural models of interacting proteins. This work revealed that disruption of physical interactions appears to be a more common disease mechanism than previously appreciated [68]. The authors found that in-frame mutations are enriched on the interaction interfaces and that mutations in different interfaces often were associated with different diseases.

Taken together, protein interaction data can play a pivotal role for translating the current flood of genetic data into a better understanding of the biological system. We therefore propose a targeted proteomic approach to unravel the molecular mechanisms behind growth or seed yield.

Author contribution

The PhD candidate wrote this chapter and designed the figure.

References

1. Alexandratos N, Bruinsma J: **World agriculture towards 2030/2050: the 2012 revision**. *ESA Work. Pap.* 2012, **No. 12-03**:1–154.

2. Castiglioni P, Warner D, Bensen RJ, Anstrom DC, Harrison J, Stoecker M, Abad M, Kumar G, Salvador S, D'Ordine R, et al.: **Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions.** *Plant Physiol.* 2008, **147**:446–455.

3. Jaggard KW, Qi A, Ober ES: **Possible changes to arable crop yields by 2050.** *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2010, **365**:2835–2851.

4. Zelazny E, Vert G: **Plant nutrition: root transporters on the move.** *Plant Physiol.* 2014, **166**:500–508.

5. Fang Z, Xia K, Yang X, Grotemeyer MS, Meier S, Rentsch D, Xu X, Zhang M: **Altered expression of the PTR/NRT1 homologue OsPTR9 affects nitrogen utilization efficiency, growth and grain yield in rice.** *Plant Biotechnol. J.* 2013, **11**:446–458.

6. Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatout C, Dubois F, Balliau T, Valot B, Davanture M, et al.: **Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production.** *Plant Cell* 2006, **18**:3252–3274.

7. Habash DZ, Massiah AJ, Rong HL, Wallsgrove RM, Leigh RA: **The role of cytosolic glutamine synthetase in wheat** *Ann. Appl. Biol.* 2001, **138**:83–89.

8. Fei H, Chaillou S, Hirel B, Mahon JD, Vessey JK: **Overexpression of a soybean cytosolic glutamine synthetase gene linked to organ-specific promoters in pea plants grown in different concentrations of nitrate.** *Planta* 2003, **216**:467–474.

9. Brauer EK, Rochon A, Bi Y-M, Bozzo GG, Rothstein SJ, Shelp BJ: **Reappraisal of nitrogen use efficiency in rice overexpressing glutamine synthetase1.** *Physiol. Plant.* 2011, **141**:361–372.

10. Shrawat AK, Carroll RT, DePauw M, Taylor GJ, Good AG: **Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of alanine aminotransferase.** *Plant Biotechnol. J.* 2008, **6**:722–732.

11. Zhu X, Long S, Ort D: **Improving photosynthetic efficiency for greater yield.** *Annu. Rev. Plant Biol.* 2010, **61**:235-261.

12. Foyer CH, Bloom AJ, Queval G, Noctor G: **Photorespiratory metabolism: genes, mutants, energetics, and redox signaling** *Annu. Rev. Plant Biol.* 2009*,* **60**: 455-484.

13. Raghavendra A, Sage R: **C4 photosynthesis and related CO2 concentrating mechanisms.** *Adv. in Photosynthesis and Respiration.* 2010.

14. Leegood RC: **Strategies for engineering C(4) photosynthesis.** *J. Plant Physiol.* 2013, **170**:378–388.

15. Hausler RE: **Overexpression of C4-cycle enzymes in transgenic C3 plants: a biotechnological approach to improve C3-photosynthesis** *J. Exp. Bot.* 2002, **53**:591–607.

16. Dohleman FG, Long SP: **More productive than maize in the Midwest: How does Miscanthus do it?** *Plant Physiol.* 2009, **150**:2104–2115.

17. Gregersen PL, Culetic A, Boschian L, Krupinska K: **Plant senescence and crop productivity.** *Plant Mol. Biol.* 2013, **82**:603–622.

18. Gan S, Amasino R: **Inhibition of leaf senescence by autoregulated production of cytokinin** *Science.* 1995, **270**: 1986-1988.

19. Guo Y, Gan S-S: **Translational researches on leaf senescence for enhancing plant productivity and quality.** *J. Exp. Bot.* 2014, **65**:3901–3913.

20. Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, et al.: **Control of rice grain-filling and yield by a gene with a potential signature of domestication.** *Nat. Genet.* 2008, **40**:1370–1374.

21. Li B, Liu H, Zhang Y, Kang T, Zhang L, Tong J, Xiao L, Zhang H: **Constitutive expression of cell wall invertase genes increases grain yield and starch content in maize.** *Plant Biotechnol. J.* 2013, **11**:1080–1091.

22. Xu S-M, Brill E, Llewellyn DJ, Furbank RT, Ruan Y-L: **Overexpression of a potato sucrose synthase gene in cotton accelerates leaf expansion, reduces seed abortion, and enhances fiber production.** *Mol. Plant* 2012, **5**:430–441.

23. Li N, Zhang S, Zhao Y, Li B, Zhang J: **Over-expression of AGPase genes enhances seed weight and starch content in transgenic maize.** *Planta* 2011, **233**:241–250.

24. Lunn JE, Feil R, Hendriks JHM, Gibon Y, Morcuende R, Osuna D, Scheible W-R, Carillo P, Hajirezaei M-R, Stitt M: **Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in Arabidopsis thaliana.** *Biochem. J.* 2006, **397**:139–148.

25. O'Hara LE, Paul MJ, Wingler A: **How do sugars regulate plant growth and development? New insight into the role of trehalose-6-phosphate.** *Mol. Plant* 2013, **6**:261–274.

26. Nunes C, O'Hara LE, Primavesi LF, Delatte TL, Schluepmann H, Somsen GW, Silva AB, Fevereiro PS, Wingler A, Paul MJ: **The trehalose 6-phosphate/SnRK1 signaling pathway primes growth recovery following relief of sink limitation.** *Plant Physiol.* 2013, **162**:1720–1732.

27. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, et al.: **Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice.** *Nat. Genet.* 2010, **42**:541–544.

28. Miura K, Ikeda M, Matsubara A, Song X-J, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M: **OsSPL14 promotes panicle branching and higher grain productivity in rice.** *Nat. Genet.* 2010, **42**:545–549.

29. Teng F, Zhai L, Liu R, Bai W, Wang L, Huo D, Tao Y, Zheng Y, Zhang Z: **ZmGA3ox2, a candidate gene for a major QTL, qPH3.1, for plant height in maize.** *Plant J.* 2013, **73**:405–416.

30. Al-Ahmad H, Dwyer J, Moloney M, Gressel J: **Mitigation of establishment of Brassica napus transgenes in volunteers using a tandem construct containing a selectively unfit gene.** *Plant Biotechnol. J.* 2006, **4**:7–21.

31. Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, Cerwick SF, Dieter JA, Duncan KE, Howard RJ, et al.: **Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield.** *J. Exp. Bot.* 2014, **65**:249–260.

32. Alonso-Blanco C, Blankestijn-de Vries H, Hanhart CJ, Koornneef M: **Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana** *Proc. Natl. Acad. Sci.* 1999, **96**:4710–4717.

33. Van Daele I, Gonzalez N, Vercauteren I, de Smet L, Inzé D, Roldán-Ruiz I, Vuylsteke M: **A comparative study of seed yield parameters in Arabidopsis thaliana mutants and transgenics.** *Plant Biotechnol. J.* 2012, **10**:488–500.

34. Gonzalez N, Beemster GTS, Inzé D: **David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops?** *Curr. Opin. Plant Biol.* 2009, **12**:157–164.

35. Vriet C, Russinova E, Reuzeau C: **Boosting crop yields with plant steroids.** *Plant Cell* 2012, **24**:842–857.

36. Huijser P, Schmid M: **The control of developmental phase transitions in plants** *Development* 2011, **138**:4117-4129.

37. Wang J-W, Czech B, Weigel D: **miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana.** *Cell* 2009, **138**:738–749.

38. Wu G, Park MY, Conway SR, Wang J-W, Weigel D, Poethig RS: **The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis.** *Cell* 2009, **138**:750–759.

39. Reuzeau C, Frankard V, Hatzfeld Y, Sanz A, Van Camp W, Lejeune P, De Wilde C, Lievens K, de Wolf J, Vranken E, et al.: **TraitMill™: a functional genomics platform for the phenotypic analysis of cereals** *Plant Genet. Resour. Charact. Util.* 2006, **4**:20–24.

40. De Bossoreille de Ribou S, Douam F, Hamant O, Frohlich MW, Negrutiu I: **Plant science and agricultural productivity: why are we hitting the yield ceiling?** *Plant Sci.* 2013, **210**:159–176.

41. Fischer R: **Can technology deliver on the yield challenge to 2050** *FAO Expert Meeting on How to Feed the World in 2050.* 2009.

42. Vanhaeren H, Gonzalez N, Coppens F, De Milde L, Van Daele T, Vermeersch M, Eloy NB, Storme V, Inzé D: **Combining growth-promoting genes leads to positive epistasis in Arabidopsis thaliana.** *Elife* 2014, **3**:e02252.

43. Vercruyssen L, Gonzalez N, Werner T, Schmülling T, Inzé D: **Combining enhanced root and shoot growth reveals cross talk between pathways that control plant organ size in Arabidopsis.** *Plant Physiol.* 2011, **155**:1339–1352.

44. Gonzalez N, De Bodt S, Sulpice R, Jikumaru Y, Chae E, Dhondt S, Van Daele T, De Milde L, Weigel D, Kamiya Y, et al.: **Increased leaf size: different means to an end.** *Plant Physiol.* 2010, **153**:1261–1279.

45. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al.: **Global mapping of the yeast genetic interaction network.** *Science* 2004, **303**:808–813.

46. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, et al.: **The genetic landscape of a cell.** *Science* 2010, **327**:425–431.

47. Horiguchi G, Kim G-T, Tsukaya H: **The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana.** *Plant J.* 2005, **43**:68–78.

48. Lara P, On L, Abraham Z, Ferra C, Dı I: **Synergistic Activation of Seed Storage Protein Gene Expression in Arabidopsis by ABI3 and Two bZIPs Related to OPAQUE2 ***. 2003, **278**:21003–21011.

49. Ham B-K, Park JM, Lee S-B, Kim MJ, Lee I-J, Kim K-J, Kwon CS, Paek K-H: **Tobacco Tsip1, a DnaJ-type Zn finger protein, is recruited to and potentiates Tsi1-mediated transcriptional activation.** *Plant Cell* 2006, **18**:2005–2020.

50. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, et al.: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437**:1173–1178.

51. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, et al.: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122**:957–968.

52. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L: **The human disease network.** *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:8685–8690.

53. Fernández E, Collins MO, Uren RT, Kopanitsa M V, Komiyama NH, Croning MDR, Zografos L, Armstrong JD, Choudhary JS, Grant SGN: **Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins.** *Mol. Syst. Biol.* 2009, **5**:269.

54. Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual J-F, Fisk CJ, Li N, Smolyar A, Hill DE, et al.: **A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration.** *Cell* 2006, **125**:801–814.

55. Camargo LM, Collura V, Rain J-C, Mizuguchi K, Hermjakob H, Kerrien S, Bonnert TP, Whiting PJ, Brandon NJ: **Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia.** *Mol. Psychiatry* 2007, **12**:74–86.

56. Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C, et al.: **A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease.** *Mol. Cell* 2004, **15**:853–865.

57. Pujana MA, Han J-DJ, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, et al.: **Network modeling links breast cancer susceptibility and centrosome dysfunction.** *Nat. Genet.* 2007, **39**:1338–1349.

58. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, et al.: **Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana.** *Mol. Syst. Biol.* 2010, **6**:397.

59. Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E: **Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network.** *Plant Cell* 2010, **22**:1264–1280.

60. Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L: **A kaleidoscopic view of the Arabidopsis core cell cycle interactome.** *Trends Plant Sci.* 2011, **16**:141–150.

61. Blomme J, Inzé D, Gonzalez N: **The cell-cycle interactome: a source of growth regulators?** *J. Exp. Bot.* 2014, **65**:2715–2730.

62. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, Wu W, Uitdehaag BMJ, Kappos L, Polman CH, et al.: **Pathway and network-based analysis of genome-wide association studies in multiple sclerosis.** *Hum. Mol. Genet.* 2009, **18**:2078–2090.

63. Barrenas F, Chavali S, Holme P, Mobini R, Benson M: **Network properties of complex human disease genes identified through genome-wide association studies.** *PLoS One* 2009, **4**:e8090.

64. Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, Ferrucci L, Tanaka T, Bandinelli S, Cho YS, et al.: **A network-based approach to prioritize results from genome-wide association studies.** *PLoS One* 2011, **6**:e24220.

65. Jia P, Zheng S, Long J, Zheng W, Zhao Z: **dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks.** *Bioinformatics* 2011, **27**:95–102.

66. Soler-López M, Zanzoni A, Lluís R, Stelzl U, Aloy P: **Interactome mapping suggests new mechanistic details underlying Alzheimer's disease.** *Genome Res.* 2011, **21**:364–376.

67. Schuster-Böckler B, Bateman A: **Protein interactions in human genetic diseases.** *Genome Biol.* 2008, **9**:R9.

68. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H: **Three-dimensional reconstruction of protein networks provides insight into human genetic disease.** *Nat. Biotechnol.* 2012, **30**:159–164.

Chapter 3: Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry

Maarten Dedecker[1,2,3], Jelle Van Leene[1,2] and Geert De Jaeger[1,2]

[1] Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium
[2] Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium
[3] CropDesign N.V., Technologiepark 21, B-9052 Ghent, Belgium

Corresponding author: De Jaeger, Geert (geert.dejaeger@psb.vib-ugent.be)

A more concise version was published as: Dedecker M, Van Leene J, De Jaeger G: **Unravelling plant molecular machineries through affinity purification coupled to mass spectrometry.** *Curr. Opin. Plant Biol.* 2015, **24**:1-9.

Research highlights

- Improved AP-MS protocols paved the way for protein complex analysis in plants
- Any sample can be assayed, from cell cultures to very specific isolated tissues
- Sensitive MS and computational filtering allow accurate detection of true interactors
- Orthogonal data integration underscores the reliability of observed interactors
- Established AP protocols allow extension to protein-RNA, -DNA, or -lipid interactions

Abstract

Rather than functioning independently, proteins tend to work in concert with each other and with other macromolecules to form macromolecular complexes. Affinity purification coupled to mass spectrometry (AP-MS) can lead to a better understanding of the cellular functions of these complexes. With the development of easy purification protocols and ultra-sensitive MS, AP-MS is currently widely used for screening co-complex membership in plants. Studying complexes in their developmental context through the isolation of specific organs and tissues has now become feasible. Besides, the tagged protein can be employed for probing other interactions like protein-DNA and protein-RNA interactions. With the tools at hand, protein-centred interaction studies will greatly improve our knowledge of how plant cells wire their functional components in relation to their function.

Introduction

The sequencing and annotation of genomes of numerous organisms revealed that the 'one-gene-one-function' hypothesis does not sufficiently explain the complexity of living organisms. Instead, proteins interact with each other and with other macromolecules to form molecular machines [1]. Currently, protein-protein interactions (PPIs) are the most rigorously studied interactions. The main techniques for screening PPIs are yeast two-hybrid (Y2H), protein complementation analysis (PCA) and affinity-purification coupled to mass spectrometry (AP-MS).

Y2H requires the co-expression of hybrid genes that encode a DNA-binding domain of the GAL4 transcription factor fused to a target protein or 'bait', and a GAL4 derived activation domain fused to a second protein, the 'prey'. Interaction of bait and prey in the yeast nucleus tether the activation domain and the DNA-binding domain, resulting in the reconstitution of a functional transcription factor that can activate a downstream reporter [2]. The problem of testing pairs of plant proteins in the yeast nucleus is twofold. First, it is a heterologous system as non-yeast proteins are tested and second, the assayed proteins are forced to the nucleus which is not necessarily their native compartment. Further, the system may lack plant-specific co-factors or post-translational modifications leading to false negative results and thus reducing sensitivity (estimated to ~36% in Arabidopsis [3]). Y2H assays are also prone to false positives due to auto-activation (e.g. in the case of transcription factors) and non-specific interactions. The latter are caused by high overexpression of proteins, incorrect protein folding or the forced localisation in the nucleus. Proteins that allow yeast to overcome nutritional selection when overexpressed are also often scored as false positives.

PCA is a more direct approach in the sense that bait and prey ORFs are fused with the two complements of a reporter gene. Interaction leads to recovery of functional reporter protein providing a measurable read-out. PCA systems are advantageous over Y2H as they 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. GFP, its derivatives and split luciferase are the most commonly used reporters for PCA in plants. The former is referred to as bimolecular fluorescence complementation (BiFC) (58). Because of the background auto-fluorescence of photosynthetic pigments in plant cells, BiFC signal detection needs to be carried out with considerable caution, making the assay not ideal for high-throughput performance [4]. The main caveat associated with BiFC is that the reporter fragments can spontaneously reassemble, especially when expression levels are strongly boosted [5], and the interactions stabilize by refolding, i.e. reconstitution of the reporter is irreversible [6]. However, the latter can sometimes be turned into an advantage because it facilitates the study of transient interactions, as has been shown for kinase substrates in plants [7]. BiFC is generally combined with transient expression in *Nicotiana benthamiana* or *Arabidopsis thaliana*, but it has been used in many other plant systems and/or with stable expression [8].

In plants, Y2H and PCA showed huge value, delivering comprehensive interaction data for both soluble [3] and membrane-linked proteins [9], and dedicated datasets ranging from for example host-pathogen interactions [10] to hormonal pathways [11,12]. As Y2H investigates interactions ectopically in yeast, PCA and AP-MS are closer to the endogenous system of plants. Also the resulting datasets differ, with PCA and Y2H delivering pairwise (direct) interactions and AP-MS yielding co-complex data (direct and indirect interactions). Therefore, the techniques tend to deliver complementary datasets [13]. Both Y2H and PCA methods are however limited in terms of the testable search space, as interactions can only be assayed from ORFs available in the screening library. The cloning step is thus a key bottleneck as the search space is restricted to the square of the number of available ORFs. For example, whereas Y2H achieves high throughput, because ORFs are available for only half of the Arabidopsis proteome, the Arabidopsis interactome mapping project is currently restricted to one-fourth of the whole interactome [3]. In contrast, affinity purification methods need only one of the interacting proteins to be cloned, and in theory these methods allow the identification of interacting proteins in the whole interactome. Practically, however, the search space is restricted to the proteins expressed in the chosen tissues.

Here, we provide an overview of the development of AP-MS in plant research, from the early adoption of the technique to cultured plant cells to the latest advancements, giving guidance on the decisions that need to be made. We consider the whole AP-MS workflow in a step-by-step fashion and show that each step is important and crucial for successful identification of interaction partners.

Strategies for expressing a tagged protein of interest

Tagging involves fusing the DNA sequence coding for an affinity portion to the open reading frame of the protein of interest and having this fusion construct expressed in a biological system. Both performance of the tag and the expression level of the construct are crucial to the success of the AP-MS experiment. Because adding an affinity handle to a protein can compromise the functionality of the protein, both N- and C-terminal fusions are preferably assessed. Despite the development and testing of a large variety of alternative tags in plants, there is a

clear overrepresentation of studies using the classical TAP tag and derivatives such as TAPi tag [14] and GS tag [15] on the one hand, and fluorescent protein tags on the other hand, reflecting their success (Figure 1). The expression level of the bait fusion construct is a second important concern. Unlike in yeast, integration of the affinity tag through homologous recombination in plants is only effective in mitochondria and plastids [16]. In the plant nucleus, the recombination efficiency drops dramatically. Therefore, the bait fusion construct is mostly introduced via *Agrobacterium tumefaciens* through random integration in the genome. Consequently, it is generally combined with either a constitutive overexpressing promoter or the endogenous promoter. Overexpression is simple and favours competition with the endogenous counterpart for incorporation in the complex. Massively overexpressed proteins may however exhibit protein misfolding, mislocalisation and/or misregulation on cellular level [17]. Functionality of the tagged protein can be verified through complementation analysis of the overexpression construct in a mutant background [18–25]. On organism level, the constitutively expressed bait protein might be present in tissues where it normally is not expressed, leading to accumulation of uncomplexed bait or even aberrant interactions. To avoid this, expression of the bait can be controlled more tightly by using synthetic inducible promoters [26]. Applicability of inducible systems in plants is however limited to cell cultured systems or hydroponics. Spraying of the inducing chemical on whole plants indeed can compromise the homogeneity of the induction. Alternatively, a 1- to 2-kb upstream sequence can be included with the bait fusion sequence to encompass the endogenous promoter. In this case, competition with the endogenous counterpart should be avoided by introducing the TAP cassette in a knock-out mutant. Still, context specific regulation of the gene of interest will be lost. A recently proposed way to circumvent this problem is to 'recombineer' the tag in the endogenous locus within a transformation-competent Bacterial Artificial Chromosome [27].

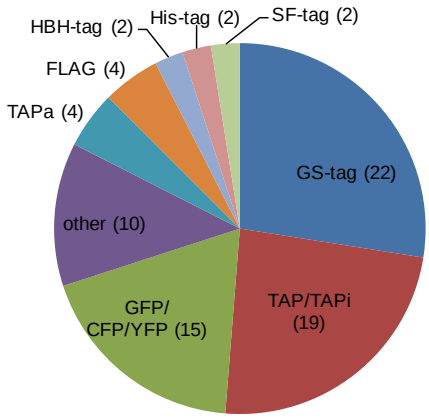
Through *Agrobacterium* mediated transformation, the T-DNA containing the TAP expression cassette is shuttled to the plant's nucleus and inserted in the genome through illegitimate recombination. Transgene individuals will thus differ mutually in insertion position and number of loci. This inevitably leads to differences in expression level of the TAP construct between individuals. Generally, this issue is addressed by selecting particular transgenic lines with similar expression levels [28,29] or by pooling multiple events [30,31]. In mammalian HEK239 cells, the Flp-In-T-Rex system allows single integration of the TAP cassette in a fixed chromosomal location [32]. This not only permits establishment of relatively homogenous cell populations, but makes the whole procedure of integrating TAP constructs into the genome more standardised. To date, there is no such system available in plant cells however.

Another promising possibility is to establish a "surrogate homologous recombination" by CRISPR-mediated insertion of affinity tags [33–36]. This could address both the issues of efficiency of incorporation in the complex and insertion locus. To our knowledge, this was not yet tested on plants however.

Figure 1. Tags successfully used for purification of protein complexes in plant AP-MS studies.

tag name	sequence/parts	purification steps	purification scheme	reference
peptide tags				
c-myc	EQKLISEEDL	single AP	α c-myc - boiling in sample buffer	suppl. [1]
FLAG	DYKDDDDK	single AP	α FLAG - FLAG peptide/boiling in sample buffer	suppl. [2]
HA	YPYDVPDYA	single AP	α-HA - pH	suppl. [3]
His-tag	HHHHHH	single AP	Ni2+-NTA - Imidazole	suppl. [4]
StrepII-tag	WSHPQFEK	single AP	StrepTactin - desthiobiotin	suppl. [5]
TEV-biotinylation sequence tomato	GTVIAPMAGLVVKVLVKDGEKVQE GQPVLVLEAMKMEHVVKAPANGYV SGLEIKVGQSVQDGIKLFALKD	single AP	streptavidin - hisTEV - Ni^{2+}-NTA	suppl. [6]
protein tags				
GFP/CFP/YFP	Fluorescent protein	single AP	α-GFP - pH/boiling in sample buffer/on-bead digestion	suppl. [7,8]
dual affinity tags				
TAP/TAPi	CBP-TEV-ProA	tandem AP single AP	IgG - TEV - calmodulin - EGTA IgG - pH	suppl. [9], [14]
TAPa	9x c-myc-6x His-HR3C-ProA	tandem AP single AP	IgG - HR3C - Ni2+-NTA - imidazole α c-myc - boiling in sample buffer	[18]
TAP2	6x His-TEV-ProA	tandem AP	IgG - TEV - GST - Ni2+-NTA - Imidazole	[73]
GS-tag	SBP-TEV-ProG	tandem AP single AP	IgG - TEV - streptavidin - desthiobiotin IgG - boiling in sample buffer	suppl. [10]
GS^{rhino}-tag	SBP-HR3C-ProG	tandem AP	IgG - HR3C - streptavidin - desthiobiotin	suppl. [11]
myc-His tag	7x myc-6x His	tandem AP	Ni2+-NTA - imidazole - α-myc - boiling in sample buffer	suppl. [12]
TAPh	CBP-TEV-ProA-6x His	tandem AP	Ni2+-NTA - imidazole - IgG - hisTEV - Ni2+-NTA	suppl. [13]
SF-tag	2x Strep-FLAG	single AP	α FLAG - FLAG	suppl. [14]
SBP-FLAG	SBP-FLAG	tandem AP	α FLAG - FLAG - streptavidin - desthiobiotin	[47]
csBP-tag	CBP-SBP	single AP	streptavidin - desthiobiotin	suppl. [15]
His-FLAG*	8x His-FLAG	single AP	Ni2+-NTA - imidazole - α-FLAG - FLAG	suppl. [16]
HA-StrepIII	3x HA-StrepIII	tandem AP	StrepTactin - desthiobiotin	suppl. [17]
HPB-tag	HA-HR3C-biotin	tandem AP	streptavidin - boiling in sample buffer	[21]
HBH-tag*	6x His-BCCD-6x His	single AP	Ni2+-NTA - imidazole - streptavidin - boiling in sample buffer	suppl. [18]

number of reports for a specific tag



Representation of the ratio of publications reporting the use of a specific tag to all plant AP-MS studies. The amount of publications for each affinity tag is shown between brackets. Tags for which the use was reported only once were grouped in the "other" category. There is a clear overrepresentation of the classic TAP tag and the more recent GS-tag. From our experience, these bigger domain-like tags tend to stabilize the expression of the bait construct and are therefore applicable on a broad variety of bait proteins. Also fluorescent proteins are frequently used. The fact that many fluorescent reporter-tagged lines are already available from localisation studies contributes to the popularity of these tags.

HA, hemagglutinin; GFP, green fluorescent protein; YFP, yellow fluorescent protein; CFP, cyan fluorescent protein; SBP, streptavidin binding peptide; CBP, calmodulin binding peptide; TEV, tobacco etch virus protease; 3CRvP, 3C human rhinovirus protease; BCCD, biotin carboxyl carrier domain. * not used for mass spectrometry

Biological systems for protein extract generation

Once expressed, the bait fusion protein provides an affinity handle to capture the complexes in which it participates. Since merely a fraction of the protein extract contains actual bait protein, reasonable amounts of total extract are required to allow identification via MS. In principle, any plant sample can be used (Figure 2), only requiring sufficient and stable bait protein complexes, but the key parameters that have an impact on the AP-MS procedure differ significantly between the different biomass sources (Figure 3).

The easiest and fastest way to screen an unknown protein is through cultured cells. Once established, they provide an unlimited supply of protein in a reasonable time with a fair workload. Suspension cells often have high ploidy levels, providing high protein content and fast growth rates. They have been used successfully for the screening of about 100 cell-cycle related proteins [37]. The system is easy to manipulate by addition or removal of chemicals (e.g. phytohormones) from the medium, enabling the focus on complex dynamics related to signalling pathways [38–40]. In suspension and callus cultures, each cell will contribute equally in relevance for obtaining complexes, requiring less sensitivity from the mass spectrometer. Cell cultures are grown isolated from any developmental context. Therefore, identified interactions give a hint about their presence in the plant, but not where or when during plant development they are actually taking place.

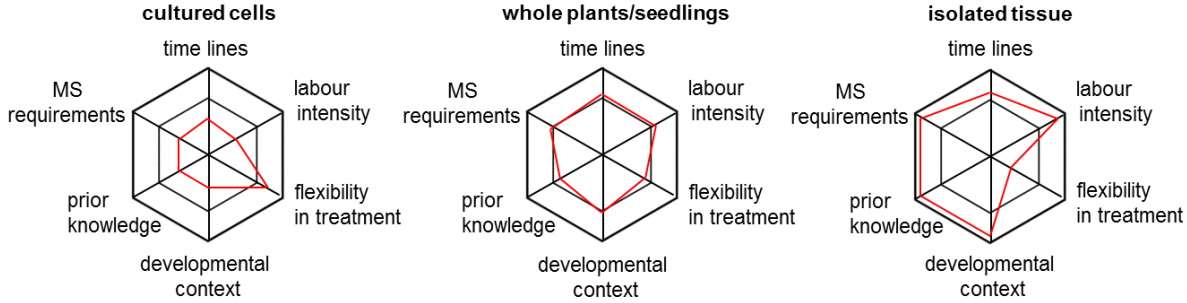
For pathways related to development, whole plants rather than cultured cells are used [41]. For example, five additional interactors of the ANGUSTIFOLIA 3 co-activator have been found *in planta* compared to cultured cells, which were confirmed through reciprocal purifications using the interactors as a bait protein [42••]. Differential analyses of protein complex dynamics in the presence or absence of environmental cues have been described [20,25,43–48]. Dependent on the tissue/cell type specificity of the bait, extraction from whole plants often leads to dilution of the relevant extract. Moreover, during extraction, the protein content from different cell types will be blended, possibly leading to false positive identifications. For baits present in the majority of cells this problem is not imminent [49,50]. Also baits functioning in dividing cells can be screened in young seedlings since these have a high ratio of propagative tissue [42••]. Thanks to advancements in sensitivity in MS, the amount of input material can be downscaled, which allows the use of plant tissues or organs instead of the whole plant, favouring the ratio of relevant tissue for more discretely expressed baits [19,21,45,47,51–54]. For instance, specifically isolating inflorescence meristems allowed Smaczniak *et al.* to sensitively detect the interaction network of floral homeotic MADS-box proteins [55••]. Tissue-specific sampling also allows to zoom in on the dynamics of low-abundance protein complexes throughout the growth zone in a developing organ [56].

Figure 2. Plant biomass sources for AP-MS.

biomass type	species/organ type	original reference *	applicable processes
cell suspension cultures 	rice Arabidopsis tobacco BY2	suppl. [19] suppl. [20] suppl. [21]	household pathways hormone signalling cell cycle
callus cultures 	rice Arabidopsis	suppl. [22] suppl. [23]	
whole plants/seedlings 	Arabidopsis	[18]	household pathways light signalling hormone signalling developmental processes stress reponses
plant organs 	rice shoots	[28]	household pathways light signalling hormone signalling developmental processes stress reponses
	Arabidopsis flowers, flower buds	[74] [55]	flowering pathways
	Arabidopsis leaves	suppl. [25]	biotic stress response plastid biogenesis/function
	Arabidopsis roots	[54]	developmental processes nutrient uptake
	tomato leaves	[47]	biotic stress response
	tobacco leaves	[24]	biotic stress response
plant tissues 	petunia pollen	[53]	self incompatibility
	growth zones of the leaf	[56]	growth/development-related processes

All types of biomass sources used for AP-MS are illustrated. Per biomass source, the first report and the applicability are represented. *References preceded by 'suppl.' are listed in the supplementary reference file.

Figure 3. Comparison on the different biomass sources for AP-MS.



Key parameters of the biomass sources that have an impact on the whole AP-MS procedure. The actual performances of each type of biomass source are compared visually by representing the individual parameters on each of the six axes. 'Time lines' and 'labour intensity' refer to how long and how much work it takes to generate enough biomass for AP-MS purification. A third parameter accounts for the 'flexibility in treatment'. This is how easily the system can be employed for dedicated experiments, such as the addition or removal of chemicals, hormones, radio-isotopes etc. With 'developmental context', we provide an estimate of how much information is acquired from the AP-MS data in the context of plant development. 'Prior knowledge' indicates how much information about the tissue specificity and developmental time frame of the bait protein is required *a priori* to maximize the success in identifying interactors from the type of biomass choses. And last, 'mass spec requirements' shows how much sensitivity is required from the mass spectrometer to detect complexes from samples.

Purification of the complex: a balance game between capturing true interactors and losing background proteins

AP-MS involves the isolation of the bait from the plant sample by affinity purification under near-physiological conditions. The bait's binding partners can then be recovered from the sample and identified by mass spectrometry.

After lysis, the cellular environment is replaced by an artificial environment with buffers, salts and stabilizers to keep complexes in solution during the purification process, while maintaining their integrity. For proteins tightly bound to chromatin or membrane proteins, the extraction protocol needs to be adapted to solubilize the targeted complexes [57]. Also, a prior fractionation step enriching for e.g. chloroplasts [16,58] or nuclei [52] can be performed for baits present in these specific subcellular fractions, increasing the sensitivity [59].

Once the cells are lysed protein complexes will start to dissociate, turning their purification in a race against the clock to capture as much interactors as possible. The bait is captured from the protein extract by adding an affinity matrix that specifically recognizes the tag. The most commonly used affinity matrices are antibodies linked to agarose, sepharose or magnetic beads. Long incubation times allow the affinity handle to reach equilibrium in binding the matrix, but also result in the dissociation of secondary interaction partners of the bait. To maximize purification of the complex the purification should be performed in a minimum amount of time. Alternatively, chemical cross-linking can be applied to enable stringent purification while preserving stable and weak or transient interactions [14,21,23]. Apart from the identities of interacting proteins, also cross-linked amino acid residues can be identified through mass spectrometry. The latter

enables direct molecular evidence describing the physical contacts between and within proteins [60]. Protein cross-linking is not frequently used however, due to inherent difficulty in the effective MS detection and accurate identification of cross-linked peptides. Most PPI studies use formaldehyde cross-linking, because of its fast kinetics and membrane permeability [14,23]. However, identification of formaldehyde cross-linked peptides is extremely challenging because of its promiscuous nonspecific reactivity and extremely short spacer length [61]. Efforts have been made to address the limitations for cross-linking studies [62], but were not transferred to plant systems yet.

The shortest purification procedures encompass one single purification step. For AP-MS experiments, the tag is selected as small as possible to minimize interference with the bait's function. Peptide tags often encompass no more than ten amino acids. The GFP-tag is also frequently used for single AP experiments, since GFP-tagged lines are already available from subcellular localisation studies. The combination of localisation and interaction data is indeed very powerful, as was demonstrated in a study identifying a bHLH heterodimer controlling vascular tissue establishment [63••]. The binding step can further be speed-up by increasing the surface area to which the tagged complex can bind. Although magnetic beads possess a 10 times smaller binding capacity, they are 100-150 times smaller than sepharose/agarose beads. Owing to their high surface area to volume, magnetic beads are therefore more suitable for short incubation times [64]. With the advancement in sensitivity of mass spectrometry, short incubation steps allowed retrieval of substrates of kinases and phosphatases in yeast [65].

Short single-step purification has the disadvantage that a considerable amount of background contaminants remains in AP-MS experiments. These often 'mask' sub-stoichiometric interactors that only appear after further purification [66,67]. Tandem purification (TAP) provides a higher signal-to-noise ratio at the cost of a lower yield. Because of the lower binding capacity of magnetic beads, they are less suitable for double-step purifications. Therefore, agarose or sepharose beads are preferentially used in these types of experiments. The TAP-tag originally developed in yeast was improved for plants, yielding the TAPi tag [14]. This tag still has limitations, being the calmodulin binding domain, which can be problematic in eukaryotic cells [15], and the TEV protease cleavage step, requiring incubation at 16°C to cleave off the Protein A domain and mildly elute the rest of the bait. Many alternatives (Figure 1), i.e. the TAPa tag [18,43,68] and the successful, highly specific GS tag [15,37], are available to deal with those weak points.

Recent increase in MS instrument power obviates the need for extensive pre-fractionation of the purified sample before MS analysis. The sample can still first be applied shortly on gel to get rid of small molecules that interfere with the mass spectrometer [41]. In some cases however, peptide extraction efficiency from gel can be poor and variable, especially in the case of hydrophobic peptides. Alternatively, gel-free alternatives were proposed. These are gaining momentum, as they provide decreased sample preparation and measurement time, and increased sensitivity [69].

Identification of 'true' interacting proteins

The generic nature of the AP experiments involves background contaminants that non-specifically bind to the tag, beads and/or complexes consistently across purifications. This characteristic is exploited by implementing control experiments, in which proteins from mock extracts are identified and integrated in a subtraction list. In that frame a "contaminant repository for affinity purification", the "CRAPome" was presented containing control pull-downs from various laboratories performed under various conditions [70]. The repository is however mainly focused on yeast and human datasets, and therefore not readily applicable for evaluating plant AP-MS experiments. Also, minor changes in the workflow can already alter the detected low abundant background binders, making the notion of a universal CRAPome problematic [69].

Different types of quantitative data from the mass spectrometer can be implemented (an extended review is found in [71]) to further resolve teasing apart *bona fide* interactors from background proteins. True interactors are expected to be enriched throughout purification compared to a control, which should be reflected in a higher measure of quantity. Although isotope labelling methods are more sensitive, they are either very expensive (Stable Isotope Labelling by Amino acids in Cell culture, [72]) or have issues with specificity (iTRAQ, [73]). Due to the autotrophy of plants, the incorporation of labelled amino acids by the plant is also less evident. Therefore, the community is more in favour of adapting label-free methods. These require high numbers of replicates with an additional statistical or probabilistic analysis for discrimination. Normalized spectral abundance factors (NSAFs) are calculated as the total number of spectra identifying each protein, normalized by the protein's length and the total number of identified spectra in the sample [74]. An alternative approach relies on peptide ion counts. Peak areas are calculated as the sum of ion counts (total ion current, TIC) over peptide elution time for all identified peptides. Relative protein concentrations are then determined by comparing peak areas of all peptides from that protein in one sample versus those in the other. A well-known algorithm using TIC is MaxQuant. A protocol based on MaxQUANT analysis of GFP- or YFP-tagged baits is already at hand [75] and has been applied in different studies [55••,63••,76••]. The method is however not ideal for analyses using experimental conditions in which controls are very clean and should therefore only be applied for single-step AP experiments [71].

Although most false positive interactions generally are discarded by simply comparing samples versus control purifications, some artificial and bait-specific interactions might remain, e.g. interactions occasionally generated during cell lysis. Therefore, in addition, proteins co-purifying at high frequency with a variety of baits can also be added to the subtraction list [41]. In fact, the large amount of unspecific binders detected over different experiments obviates the use of a classic untagged control strain and enables comparing to a control group consisting of many unrelated pull-downs instead [69]. Correlation of the intensities of the potential interactor with the profile of the bait protein over different experiments could help in further fine-tuning classification of interactors from background [69].

Alternatively, different scoring methods were developed to estimate the reliability of individual associations in interaction datasets from yeast and animal cells [77–82]. These estimates are then benchmarked against a reference set of reliable, known interactions – the 'Gold Standard' – and used to derive a final high confidence network containing only PPI of acceptable confidence level. However, these scoring are usually developed (and optimised) for a particular AP-MS dataset. A recent comparison of six popular scoring methods for six different types of datasets showed substantial variation in performance of the scoring method depending on the dataset [83]. More disturbingly, there was very poor overlap between the high confidence datasets from the 6 methods when applied on a same AP-MS dataset [83]. The high confidence networks resulting from each scoring method were enriched in high abundance proteins and depleted in low abundance ones, suggesting that the scoring schemes used to build them may not be as effective as expected. The same bias was also present in the reference sets or Golden Standards [83]. These factors will need to be addressed by the field to move forward.

Validating and interpreting the list of potential interaction partners

Validation can be achieved through confirmation in biological repeat experiments or through 'reciprocal tagging', by performing additional AP-MS experiments using the interaction partners as bait. This was illustrated by the elucidation of the TPLATE complex, in which an eight-subunit core complex was unravelled through extensive screening of all initial interaction partners. Furthermore, the interaction network was as such extended, identifying linkage of this core complex to the plant endocytosis machinery [76••]. Alternatively, interactors have been validated with orthogonal PPI approaches such as Y2H [29,84•], split-luciferase [37] or bimolecular fluorescence complementation [55••,85]. Using GFP as tag, its fluorescent characteristics can be applied to confirm interactions with FRET/FLIM analysis [63••]. Validation by pairwise interaction screening methods can lead to considerable amounts of false negatives, because the overlap with affinity purified complex data is inherently low [13]. This is partly due to AP-MS experiments interrogating a different subspace within the whole interactome than Y2H and PCA, and partly because PPI screening methods are associated with low sensitivity. Extra confidence in identified prey proteins can be obtained by looking for enrichment in relevant features, related to the pathway under investigation [37]. On top, the transcript expression correlation between interactors can be verified as a measure of confidence [86], together with other orthogonal data as protein expression data and/or co-localisation data. An overview of available datasets is shown in Table 1.

The output of an AP-MS experiment is still far away from biological reality, since merely a list of potential interaction partners is generated. Additional binary PPI approaches can help in unravelling the co-complex data into direct interactions. Also co-expressed sub-members of the complex will be more likely involved in a direct interaction. Additional literature searches can further help in adding knowledge to unravel the biological relevance [85]. Smaller scale studies use the outcome of AP-MS experiments as a starting point for thorough molecular study. In these cases, knockout or overexpression mutants of preys are analysed and

compared with mutant phenotypes of the bait. Also (co-) localisation studies can be applied, especially if the bait was tagged with a fluorescent moiety.

Table 1. Repositories for orthogonal data to strengthen confidence in identified interactions.

data type	rationale	example repository	reference
co-expression data	Interactors with strong expression correlation are often part of a common molecular assembly.	ATTED-II BAR CorNet	suppl. [26] suppl. [27] suppl. [28]
protein expression data	Prerequisite for interaction is that both proteins are present in the same tissue/organ at the same time.	GATOR	suppl. [29]
(sub-)localisation data	Prerequisite for interaction is that both proteins are present in the same compartment at the same time.	SUBA	suppl. [30]
annotation data	Interacting proteins are expected to function in a common pathway.	PlantGO GOMapMan Arabidopsis Reactome	suppl. [31] suppl. [32] suppl. [33]
phenotypic data	Mutants for interacting proteins showing a similar phenotype increases the confidence that both proteins co-operate.	RARGE II RAPID RMD	suppl. [34] suppl. [35] suppl. [36]
protein-protein interaction data	Identified interactions might overlap with previously detected or predicted interactions.	AtPIN AtPID BAR, CorNet	suppl. [37] suppl. [38] suppl. [39,40]
link in literature	An earlier hypothesized link between interacting proteins can further strengthen confidence in interactions.	EVEX Textpresso	suppl. [41] suppl. [42]

*References preceded by 'suppl.' are listed in the supplementary reference file.

Broadening the applicability of tagged proteins

Proteins of course do not only interact with other proteins in a cellular environment. Transcription factors target specific binding elements on DNA to recruit or block the transcription machinery. Chromatin immune-precipitation (ChIP) is used for the identification of genomic regions where specific proteins are associated. The use of antibodies can be omitted by tagging the DNA binding protein, and performing the appropriate affinity purification [48]. More recently a tandem chromatin affinity purification or TChAP protocol has been proposed [87•], improving the DNA enrichment ratios. Combined with Illumina sequencing, it enables the identification of novel protein-DNA interactions. The downside is that more material is required, and therefore the method was currently only used on cultured cells [42••,88]. A TChAP approach can also be applied for studying protein-RNA interactions, as has been illustrated by the elucidation of the interactions of ARGONAUTE 10 (AGO10) with specific miRNAs, thereby acting as a decoy for AGO1 and maintaining shoot apical meristem identity [89•]. Another variation using tagged proteins allowed the study of protein-lipid interactions for yeast proteins [90]. Alternatively, the TAP system was utilized in an innovative way to trap and identify sulfenylated proteins in H_2O_2 stress responses in plant

cells [91•]. This shows that an established AP-MS protocol can further be extended in exploring interactions of the bait protein with all other kinds of biomolecules.

Instead of a targeted search for interaction partners using proteins cloned in appropriate expression cassettes, an alternative method has been proposed recently relying on immediate chromatographic separation and MS analysis of extracts from subcellular fractions of plant tissues [92]. This method has the potential to give a comprehensive read-out of the native complexes present in the sample tested.

Concluding remarks

The dawn of major technical improvements in protein complex purification, mass spectrometry and bioinformatics paved the way for facilitated analysis of protein interactions. The increase in sensitivity of the protocols permits to zoom in on specific plant tissues, enabling to follow complex compositions in a developmental context. This will allow to estimate the dynamics of protein complex (dis)assembly during plant growth and to get a better understanding of how protein interactions play a role in important plant growth processes.

Acknowledgements

Maarten Dedecker is supported by a predoctoral fellowship from the Agency for Innovation by Science and Technology and CropDesign N.V. (‘Baekeland’ funding, IWT 100222). Jelle Van Leene is a Postdoctoral Fellow of the Research Foundation-Flanders. The authors thank Annick Bleys for help in preparing the manuscript.

Author contribution

The PhD candidate wrote this review and designed the figures, JVL and GDJ commented on the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

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

2. Fields S, Song O: **A novel genetic system to detect protein-protein interactions.** *Nature* 1989, **340**:245–246.

3. **Evidence for network evolution in an Arabidopsis interactome map.** *Science* 2011, **333**:601–607.

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

5. Zamyatnin AA, Solovyev AG, Bozhkov P V, Valkonen JPT, Morozov SY, Savenkov EI: **Assessment of the integral membrane protein topology in living cells.** *Plant J.* 2006, **46**:145–154.

6. Magliery TJ, Wilson CGM, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L: **Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism.** *J. Am. Chem. Soc.* 2005, **127**:146–157.

7. Pusch S, Harashima H, Schnittger A: **Identification of kinase substrates by bimolecular complementation assays.** *Plant J.* 2012, **70**:348–356.

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

9. Jones AM, Xuan Y, Xu M, Wang R-S, Ho C-H, Lalonde S, You CH, Sardi MI, Parsa S a, Smith-Valle E, et al.: **Border control--a membrane-linked interactome of Arabidopsis.** *Science* 2014, **344**:711–716.

10. Mukhtar MS, Carvunis A-R, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, et al.: **Independently evolved virulence effectors converge onto hubs in a plant immune system network.** *Science* 2011, **333**:596–601.

11. Lumba S, Toh S, Handfield L-F, Swan M, Liu R, Youn J-Y, Cutler SR, Subramaniam R, Provart N, Moses A, et al.: **A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in Arabidopsis.** *Dev. Cell* 2014, **29**:360–372.

12. Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, et al.: **The auxin signalling network translates dynamic input into robust patterning at the shoot apex.** *Mol. Syst. Biol.* 2011, **7**:508.

13. Yu H, Braun P, Yildirim M a, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al.: **High-quality binary protein interaction map of the yeast interactome network.** *Science* 2008, **322**:104–110.

14. Rohila JS, Chen M, Cerny R, Fromm ME: **Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants.** *Plant J.* 2004, **38**:172–181.

15. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

16. Suzuki JY, Ytterberg a J, Beardslee T a, Allison L a, Wijk KJ, Maliga P: **Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme.** *Plant J.* 2004, **40**:164–172.

17. Gibson TJ, Seiler M, Veitia RA: **The transience of transient overexpression.** *Nat. Methods* 2013, **10**:715–721.

18. Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, Dinesh-Kumar SP, Deng XW: **An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation.** *Plant J.* 2005, **41**:767–778.

19. Xing D, Chen Z: **Effects of mutations and constitutive overexpression of EDS1 and PAD4 on plant resistance to different types of microbial pathogens.** *Plant Sci.* 2006, **171**:251–262.

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

21. Qi Y, Katagiri F: **Purification of low-abundance Arabidopsis plasma-membrane protein complexes and identification of candidate components.** *Plant J.* 2009, **57**:932–944.

22. Nishimura N, Sarkeshik A, Nito K, Park S-Y, Wang A, Carvalho PC, Lee S, Caddell DF, Cutler SR, Chory J, et al.: **PYR/PYL/RCAR family members are major in-vivo ABI1 protein phosphatase 2C-interacting proteins in Arabidopsis.** *Plant J.* 2010, **61**:290–299.

23. Golisz A, Sikorski PJ, Kruszka K, Kufel J: **Arabidopsis thaliana LSM proteins function in mRNA splicing and degradation.** *Nucleic Acids Res.* 2013, **41**:6232–6249.

24. Dong H, Fei G-L, Wu C-Y, Wu F-Q, Sun Y-Y, Chen M-J, Ren Y-L, Zhou K-N, Cheng Z-J, Wang J-L, et al.: **A rice virescent-yellow leaf mutant reveals new insights into the role and assembly of plastid caseinolytic protease in higher plants.** *Plant Physiol.* 2013, **162**:1867–1880.

25. Choi H, Jeong S, Kim DS, Na HJ, Ryu JS, Lee SS, Nam HG, Lim PO, Woo HR: **The homeodomain-leucine zipper ATHB23, a phytochrome B-interacting protein, is important for phytochrome B-mediated red light signaling.** *Physiol. Plant.* 2014, **150**:308–320.

26. Nishikiori M, Mori M, Dohi K, Okamura H, Katoh E, Naito S, Meshi T, Ishikawa M: **A host small GTP-binding protein ARL8 plays crucial roles in tobamovirus RNA replication.** *PLoS Pathog.* 2011, **7**:e1002409.

27. Zhou R, Benavente LM, Stepanova AN, Alonso JM: **A recombineering-based gene tagging system for Arabidopsis.** *Plant J.* 2011, **66**:712–723.

28. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

29. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

30. Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, et al.: **A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana.** *Mol. Cell. Proteomics* 2007, **6**:1226–1238.

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

32. Glatter T, Wepf A, Aebersold R, Gstaiger M: **An integrated workflow for charting the human interaction proteome: insights into the PP2A system.** *Mol. Syst. Biol.* 2009, **5**:237.

33. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al.: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339**:819–823.

34. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339**:823–826.

35. Park A, Won ST, Pentecost M, Bartkowski W, Lee B: **CRISPR/Cas9 allows efficient and complete knock-in of a destabilization domain-tagged**

essential protein in a human cell line, allowing rapid knockdown of protein function. *PLoS One* 2014, **9**:e95101.

36. Böttcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, Wieland I, Gaul U, Förstemann K: **Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells.** *Nucleic Acids Res.* 2014, **42**:e89.

37. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, et al.: **Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana.** *Mol. Syst. Biol.* 2010, **6**:397.

38. Antoni R, Gonzalez-Guzman M, Rodriguez L, Peirats-Llobet M, Pizzio G a, Fernandez M a, De Winne N, De Jaeger G, Dietrich D, Bennett MJ, et al.: **PYRABACTIN RESISTANCE1-LIKE8 plays an important role for the regulation of abscisic acid signaling in root.** *Plant Physiol.* 2013, **161**:931–941.

39. Bassard J-E, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, et al.: **Protein-protein and protein-membrane associations in the lignin pathway.** *Plant Cell* 2012, **24**:4465–4482.

40. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Pérez AC, Chico JM, Bossche R Vanden, Sewell J, Gil E, et al.: **NINJA connects the co-repressor TOPLESS to jasmonate signalling.** *Nature* 2010, **464**:788–791.

41. Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, et al.: **An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes.** *Nat. Protoc.* 2015, **10**:169–187.

42. Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han S-K, Jégu T, Archacki R, Van Leene J, Andriankaja M, et al.: **ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development.** *Plant Cell* 2014, **26**:210–229.

•• The authors identified ANGUSTAFOLIA3 as an interacting protein of SWI/SNF complexes in Arabidopsis and complemented differential expression data from inducible nuclear localisation of AN3 with tandem chromatin affinity purification coupled to sequencing (TChAP-seq) and TAP-MS data to show that AN3 associates with chromatin remodelers to regulate transcription required for transition from cell proliferation to cell differentiation in the developing leaf.

43. Saijo Y, Zhu D, Li J, Rubio V, Zhou Z, Shen Y, Hoecker U, Wang H, Deng XW: **Arabidopsis COP1/SPA1 complex and FHY1/FHY3 associate with distinct phosphorylated forms of phytochrome A in balancing light signaling.** *Mol. Cell* 2008, **31**:607–613.

44. Phee B-K, Shin DH, Cho J-H, Kim S-H, Kim J-I, Lee Y-H, Jeon J-S, Bhoo SH, Hahn T-R: **Identification of phytochrome-interacting protein candidates in Arabidopsis thaliana by co-immunoprecipitation coupled with MALDI-TOF MS.** *Proteomics* 2006, **6**:3671–3680.

45. Liebrand TWH, Smit P, Abd-El-Haliem A, de Jonge R, Cordewener JHG, America AHP, Sklenar J, Jones AME, Robatzek S, Thomma BPHJ, et al.: **Endoplasmic reticulum-quality control chaperones facilitate the biogenesis of Cf receptor-like proteins involved in pathogen resistance of tomato.** *Plant Physiol.* 2012, **159**:1819–1833.

46. Kaiserli E, Sullivan S, Jones M a, Feeney K a, Christie JM: **Domain swapping to assess the mechanistic basis of Arabidopsis phototropin 1 receptor kinase activation and endocytosis by blue light.** *Plant Cell* 2009, **21**:3226–3244.

47. Gutierrez JR, Balmuth AL, Ntoukakis V, Mucyn TS, Gimenez-Ibanez S, Jones AME, Rathjen JP: **Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition.** *Plant J.* 2010, **61**:507–518.

48. De Lucia F, Crevillen P, Jones AME, Greb T, Dean C: **A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:16831–16836.

49. Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I: **Identification and characterization of nuclear pore complex components in Arabidopsis thaliana.** *Plant Cell* 2010, **22**:4084–4097.

50. Book AJ, Gladman NP, Lee S-S, Scalf M, Smith LM, Vierstra RD: **Affinity purification of the Arabidopsis 26 S proteasome reveals a diverse array of plant proteolytic complexes.** *J. Biol. Chem.* 2010, **285**:25554–25569.

51. Tameling WIL, Baulcombe DC: **Physical association of the NB-LRR resistance protein Rx with a Ran GTPase-activating protein is required for extreme resistance to Potato virus X.** *Plant Cell* 2007, **19**:1682–1694.

52. Monaghan J, Xu F, Gao M, Zhao Q, Palma K, Long C, Chen S, Zhang Y, Li X: **Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity.** *PLoS Pathog.* 2009, **5**:e1000526.

53. Li S, Sun P, Williams JS, Kao T: **Identification of the self-incompatibility locus F-box protein-containing complex in Petunia inflata.** *Plant Reprod.* 2014, **27**:31–45.

54. Fujiwara M, Uemura T, Ebine K, Nishimori Y, Ueda T, Nakano A, Sato MH, Fukao Y: **Interactomics of Qa-SNARE in Arabidopsis thaliana.** *Plant Cell Physiol.* 2014, **55**:781–789.

55. Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, Dinh QDP, Liu S, Westphal AH, Boeren S, et al.: **Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:1560–1565.

•• This paper describes the experimental validation that the five major floral homeotic MADS domain proteins interact in floral tissues as proposed in the 'floral quartet' model. AP-MS experiments were performed on GFP- or YFP-tagged MADS box proteins, and combined with MaxQUANT quantitative data analyses. These showed that MADS box proteins not only mutually interact, but also associate with other types of transcriptional regulators in Arabidopsis plants.

56. Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, et al.: **Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf.** *Plant Cell* 2015, **27**:1605–1619.

57. Babu M, Vlasblom J, Pu S, Guo X, Graham C, Bean BDM, Burston HE, Vizeacoumar FJ, Snider J, Phanse S, et al.: **Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae.** *Nature* 2012, **489**:585–589.

58. Stoppel R, Manavski N, Schein A, Schuster G, Teubner M, Schmitz-Linneweber C, Meurer J: **RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the Arabidopsis chloroplast.** *Nucleic Acids Res.* 2012, **40**:8593–8606.

59. Lavallée-Adam M, Rousseau J, Domecq C, Bouchard A, Forget D, Faubert D, Blanchette M, Coulombe B: **Discovery of cell compartment specific protein-protein interactions using affinity purification combined with tandem mass spectrometry.** *J. Proteome Res.* 2013, **12**:272–281.

60. Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R: **Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics.** *Mol. Cell. Proteomics* 2010, **9**:1634–1649.

61. Sutherland BW, Toews J, Kast J: **Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions.** *J. Mass Spectrom*. 2008, **43**:699–715.

62. Kaake RM, Wang X, Burke A, Yu C, Kandur W, Yang Y, Novtisky EJ, Second T, Duan J, Kao A, et al.: **A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells.** *Mol. Cell. Proteomics* 2014, **13**:3533–3543.

63. De Rybel B, Möller B, Yoshida S, Grabowicz I, Barbier de Reuille P, Boeren S, Smith RS, Borst JW, Weijers D: **A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis.** *Dev. Cell* 2013, **24**:426–437.

•• An auxin-regulated basic helix-loop-helix (bHLH) transcription factor dimer was identified as a critical regulator of vascular development. The dimer formation was uncovered by AP-MS using GFP and MaxQUANT data analyses. The dimer was shown to operate independently of tissue identity, but the overlap of transcription patterns for the interacting bHLH subunits was restricted to a small vascular domain. The study offers a framework for developmental control of local cell divisions.

64. Hubner N, Mann M: **Extracting gene function from protein–protein interactions using Quantitative BAC InteraCtomics (QUBIC).** *Methods* 2011, **53**: 453-459.

65. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin Z-Y, Breitkreutz B-J, Stark C, Liu G, et al.: **A global protein kinase and phosphatase interaction network in yeast.** *Science* 2010, **328**:1043–1046.

66. Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH: **A protein interaction network for pluripotency of embryonic stem cells.** *Nature* 2006, **444**:364–368.

67. Fernández E, Collins MO, Uren RT, Kopanitsa M V, Komiyama NH, Croning MDR, Zografos L, Armstrong JD, Choudhary JS, Grant SGN: **Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins.** *Mol. Syst. Biol.* 2009, **5**:269.

68. Templeton GW, Nimick M, Morrice N, Campbell D, Goudreault M, Gingras A-C, Takemiya A, Shimazaki K-I, Moorhead GBG: **Identification and characterization of AtI-2, an Arabidopsis homologue of an ancient protein phosphatase 1 (PP1) regulatory subunit.** *Biochem. J.* 2011, **435**:73–83.

69. Keilhauer EC, Hein MY, Mann M: **Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS).** *Mol. Cell. Proteomics* 2015, **14**:120–135.

70. Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva Y V, Hauri S, Sardiu ME, Low TY, et al.: **The CRAPome: a**

contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* 2013, **10**:730–736.

71. Pardo M, Choudhary JS: **Assignment of protein interactions from affinity purification/mass spectrometry data.** *J. Proteome Res.* 2012, **11**:1462–1474.

72. Rodrigues RB, Sabat G, Minkoff BB, Burch HL, Nguyen TT, Sussman MR: **Expression of a translationally fused TAP-tagged plasma membrane proton pump in Arabidopsis thaliana.** *Biochemistry* 2014, **53**:566–578.

73. Chang I-F, Curran A, Woolsey R, Quilici D, Cushman JC, Mittler R, Harmon A, Harper JF: **Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in Arabidopsis thaliana.** *Proteomics* 2009, **9**:2967–2985.

74. Law J a, Ausin I, Johnson LM, Vashisht A a, Zhu J-K, Wohlschlegel J a, Jacobsen SE: **A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis.** *Curr. Biol.* 2010, **20**:951–956.

75. Smaczniak C, Li N, Boeren S, America T, van Dongen W, Goerdayal SS, de Vries S, Angenent GC, Kaufmann K: **Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues.** *Nat. Protoc.* 2012, **7**:2144–2158.

76. Gadeyne A, Sánchez-Rodríguez C, Vanneste S, Di Rubbo S, Zauber H, Vanneste K, Van Leene J, De Winne N, Eeckhout D, Persiau G, et al.: **The TPLATE adaptor complex drives clathrin-mediated endocytosis in plants.** *Cell* 2014, **156**:691–704.

•• The authors identified the TPLATE complex, essential for plant growth via its role as major adaptor module for clathrin-mediated endocytosis. The complex was unravelled through exhaustive (reciprocal) TAP and single AP purifications of the different subunits. Validation of the complex was achieved with binary PPI approaches, and further confirmation was obtained through shared mutant phenotypes.

77. Choi H, Larsen B, Lin Z-Y, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras A-C, Nesvizhskii AI: **SAINT: probabilistic scoring of affinity purification-mass spectrometry data.** *Nat. Methods* 2011, **8**:70–73.

78. Sowa ME, Bennett EJ, Gygi SP, Harper JW: **Defining the human deubiquitinating enzyme interaction landscape.** *Cell* 2009, **138**:389–403.

79. Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, et al.: **A protein complex network of Drosophila melanogaster.** *Cell* 2011, **147**:690–703.

80. Collins SR, Kemmeren P, Zhao X-C, Greenblatt JF, Spencer F, Holstege FCP, Weissman JS, Krogan NJ: **Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae.** *Mol. Cell. Proteomics* 2007, **6**:439–450.

81. Hart GT, Lee I, Marcotte ER: **A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality.** *BMC Bioinformatics* 2007, **8**:236.

82. Zhang B, Park B, Karpinets T, Samatova NF: **From pull-down data to protein interaction networks and complexes with biological relevance**. *Bioinformatics* 2008, **24**: 979-986.

83. Pu S, Vlasblom J, Turinsky A, Marcon E, Phanse S, Trimble SS, Olsen J, Greenblatt J, Emili A, Wodak SJ: **Extracting high confidence protein interactions from affinity purification data: at the crossroads.** *J. Proteomics* 2015, **118**:63–80.

84. Eloy NB, Gonzalez N, Van Leene J, Maleux K, Vanhaeren H, De Milde L, Dhondt S, Vercruysse L, Witters E, Mercier R, et al.: **SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:13853–13858.

• In this paper, the authors describe the characterisation of SAMBA, a plant specific negative regulator of the APC/C complex. The interaction between SAMBA and the APC/C complex, initially detected in the screening of the cell-cycle interactome, was confirmed in planta using TAP on seedlings. Y2H assays with the SAMBA protein against all APC subunits and its co-activators provided insight in the direct interactions.

85. Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L: **A kaleidoscopic view of the Arabidopsis core cell cycle interactome.** *Trends Plant Sci.* 2011, **16**:141–150.

86. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han J-DJ, Hao T, Berriz GF, Bertin N, Huang J, Chuang L-S, et al.: **Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis.** *Nature* 2005, **436**:861–865.

87. Verkest A, Abeel T, Heyndrickx KS, Van Leene J, Lanz C, Van De Slijke E, De Winne N, Eeckhout D, Persiau G, Van Breusegem F, et al.: **A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification.** *Plant Physiol.* 2014, **164**:1122–1133.

• This is a proof of concept paper about tandem chromatin affinity purification (TChAP). The method was evaluated by comparing TChAP with ChIP and single-step ChAP in Arabidopsis cell suspension cultures. The TChAP method showed to improve DNA enrichment ratios compared

to the single affinity approaches at the cost of required initial cell material.

88. Heyman J, Cools T, Vandenbussche F, Heyndrickx KS, Van Leene J, Vercauteren I, Vanderauwera S, Vandepoele K, De Jaeger G, Van Der Straeten D, et al.: **ERF115 controls root quiescent center cell division and stem cell replenishment.** *Science* 2013, **342**:860–863.

89. Zhu H, Hu F, Wang R, Zhou X, Sze S-H, Liou LW, Barefoot A, Dickman M, Zhang X: **Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development.** *Cell* 2011, **145**:242–256.

• The authors show the first application of screening for protein-RNA interactions in plants. Small noncoding RNAs bound to AGO10 were tandem affinity purified, cloned and identified using Illumina sequencing. The authors were able to show that AGO10 specifically interacts with miR166/165, functioning as a decoy to maintain the shoot apical meristem.

90. Maeda K, Poletto M, Chiapparino A, Gavin A-C: **A generic protocol for the purification and characterization of water-soluble complexes of affinity-tagged proteins and lipids.** *Nat. Protoc.* 2014, **9**:2256–2266.

91. Waszczak C, Akter S, Eeckhout D, Persiau G, Wahni K, Bodra N, Van Molle I, De Smet B, Vertommen D, Gevaert K, et al.: **Sulfenome mining in Arabidopsis thaliana.** *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:11545–11550.

• The authors present a trapping system for sulfenylated proteins based on a heterologous yeast transcription factor, C-terminally fused to a tandem affinity purification tag. The performance of the system was illustrated with the identification of ~100 sulfenylated proteins in Arabidopsis cultured cells exposed to H_2O_2 stress.

92. Aryal UK, Xiong Y, McBride Z, Kihara D, Xie J, Hall MC, Szymanski DB: **A Proteomic Strategy for Global Analysis of Plant Protein Complexes.** *Plant Cell* 2014, 26: 3867-3882.

Supplementary references:

Supplementary references to Figure 1: suppl. [1-18]

Supplementary references to Figure 2: suppl. [19-25]

Supplementary references to Table 1: suppl. [26-42]

1. Evan GI, Hancock DC: **Studies on the interaction of the human c-*myc* protein with cell nuclei: p62$^{c\text{-}myc}$ as a member of a discrete subset of nuclear proteins.** *Cell* 1985, **43**:253-261.

2. Hopp TP, Prickett KS, Price VL, Libby RT, March CJ, Pat Cerretti D, Urdal DL, Conlon PJ: **A short polypeptide marker sequence useful for recombinant protein identification and purification.** *Nat. Biotechnol.* 1988, **6**:1204-1210.

3. Field J, Nikawa J-I, Broek D, MacDonald B, Rodgers L, Wilson IA, Lerner RA, Wigler M: **Purification of a *RAS*-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method.** *Mol. Cell. Biol.* 1988, **8**:2159-2165.

4. Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D: **Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent.** *Nat. Biotechnol.* 1988, **6**:1321-1325.

5. Schmidt TGM, Skerra A: **One-step affinity purification of bacterially produced proteins by means of the 'Strep tag' and immobilized recombinant core streptavidin.** *J. Chromatogr.* 1994, **676**:337-345.

6. de Boer AH, van Hunnik E, Korthout HAAJ, Sedee NJA, Wang M: **Affinity purification of GTPase proteins from oat root plasma membranes using biotinylated GTP.** *FEBS Lett.* 1994, **337**:281-284.

7. Lorković ZJ, Hilscher J, Barta A: **Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells.** *Mol. Biol. Cell* 2004, **15**:3233-3243.

8. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H: **A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins.** *Mol. Cell. Proteomics* 2008, **7**:282-289.

9. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B: **A generic protein purification method for protein complex characterization and proteome exploration.** *Nat. Biotechnol.* 1999, **17**:1030-1032.

10. Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superti-Furga G, Bauch A: **An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells.** *Nat. Methods* 2006, **3**:1013-1019.

11. Heijde M, Binkert M, Yin R, Ares-Orpel F, Rizzini L, Van De Slijke E, Persiau G, Nolf J, Gevaert K, De Jaeger G *et al.*: **Constitutively active UVR8 photoreceptor variant in *Arabidopsis*.** *Proc. Natl. Acad. Sci. U.S.A.* 2013, **110**:20326-20331.

12. Graumann J, Dunipace LA, Seol JH, McDonald WH, Yates JR, Wold BJ, Deshaies RJ: **Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast.** *Mol. Cell. Proteomics* 2004, **3**:226-237.

13. Wang C, Shang J-X, Chen Q-X, Oses-Prieto Ja, Bai M-Y, Yang Y, Yuan M, Zhang Y-L, Mu C-C, Deng Z *et al.*: **Identification of BZR1-interacting proteins as potential components of the brassinosteroid signaling**

pathway in *Arabidopsis* through tandem affinity purification. *Mol. Cell. Proteomics* 2013, **12**:3653-3665.

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

15. Gustafson MP, Welcker M, Hwang HC, Clurman BE: **Zcchc8 is a glycogen synthase kinase-3 substrate that interacts with RNA-binding proteins.** *Biochem. Biophys. Res. Commun.* 2005, **338**:1359-1367.

16. Kaneko A, Umeyama T, Hanaoka N, Monk BC, Uehara Y, Niimi M: **Tandem affinity purification of the *Candida albicans* septin protein complex.** *Yeast* 2004, **21**:1025-1033.

17. Stoppel R, Manavski N, Schein A, Schuster G, Teubner M, Schmitz-Linneweber C, Meurer J: **RHON1 is a novel ribonucleic acid-binding protein that supports RNase E function in the *Arabidopsis* chloroplast.** *Nucleic. Acids Res.* 2012, **40**:8593-8606.

18. Tagwerker C, Zhang H, Wang X, Larsen LSZ, Lathrop RH, Hatfield GW, Auer B, Huang L, Kaiser P: **HB tag modules for PCR-based gene tagging and tandem affinity purification in *Saccharomyces cerevisiae*.** *Yeast* 2006, **23**:623-632.

19. Zhong J, Haynes PA, Zhang S, Yang X, Andon NL, Eckert D, Yates JR, Wang X, Budworth P: **Development of a system for the study of protein-protein interactions *in planta*: characterization of a TATA-box binding protein complex in *Oryza sativa*.** *J. Proteome Res.* 2003, **2**:514-522.

20. Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N *et al.*: **A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*.** *Mol. Cell. Proteomics* 2007, **6**:1226-1238.

21. Nishikiori M, Mori M, Dohi K, Okamura H, Katoh E, Naito S, Meshi T, Ishikawa M: **A host small GTP-binding protein ARL8 plays crucial roles in tobamovirus RNA replication.** *PLoS Pathog.* 2011, **7**:e1002409.

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

23. Jeong IS, Fukudome A, Aksoy E, Bang WY, Kim S, Guan Q, Bahk JD, May Ka, Russell WK, Zhu J *et al.*: **Regulation of abiotic stress signalling by Arabidopsis C-terminal domain phosphatase-like 1 requires interaction with a K-homology domain-containing protein.** *PLoS ONE* 2013, **8**:e80509.

24. Rohila JS, Chen M, Cerny R, Fromm ME: **Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants.** *Plant J.* 2004, **38**:172-181.

25. Witte C-P, Noël LD, Gielbert J, Parker JE, Romeis T: **Rapid one-step protein purification from plant material using the eight-amino acid StrepII epitope.** *Plant Mol. Biol.* 2004, **55**:135-147.

26. Obayashi T, Okamura Y, Ito S, Tadaka S, Aoki Y, Shirota M, Kinoshita K: **ATTED-II in 2014: evaluation of gene coexpression in agriculturally important plants.** *Plant Cell Physiol.* 2014, **55**:e6.

27. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ: **An 'electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets.** *PLoS ONE* 2007, **2**:e718.

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

29. Joshi HJ, Hirsch-Hoffmann M, Baerenfaller K, Gruissem W, Baginsky S, Schmidt R, Schulze WX, Sun Q, van Wijk KJ, Egelhofer V *et al.*: **MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data.** *Plant Physiol.* 2011, **155**:259-270.

30. Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH: **SUBA: the Arabidopsis subcellular database.** *Nucleic Acids Res.* 2007, **35**:D213-218.

31. The Plant Ontology Consortium: **The Plant Ontology™ Consortium and plant ontologies.** *Comp. Funct. Genomics* 2002, **3**:137-142.

32. Ramšak Ž, Baebler Š, Rotter A, Korbar M, Mozetič I, Usadel B, Gruden K: **GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology.** *Nucleic Acids Res.* 2014, **42**:D1167-1175.

33. Tsesmetzis N, Couchman M, Higgins J, Smith A, Doonan JH, Seifert GJ, Schmidt EE, Vastrik I, Birney E, Wu G *et al.*: **Arabidopsis reactome: a foundation knowledgebase for plant systems biology.** *Plant Cell* 2008, **20**:1426-1436.

34. Akiyama K, Kurotani A, Iida K, Kuromori T, Shinozaki K, Sakurai T: **RARGE II: an integrated phenotype database of Arabidopsis mutant traits using a controlled vocabulary.** *Plant Cell Physiol.* 2014, **55**:e4.

35. Kuromori T, Wada T, Kamiya A, Yuguchi M, Yokouchi T, Imura Y, Takabe H, Sakurai T, Akiyama K, Hirayama T *et al.*: **A trial of phenome analysis using 4000 *Ds*-insertional mutants in gene-coding regions of Arabidopsis.** *Plant J.* 2006, **47**:640-651.

36. Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q, Wang S: **RMD: a rice mutant database for functional analysis of the rice genome.** *Nucleic Acids Res.* 2006, **34**:D745-748.

37. Brandão MM, Dantas LL, Silva-Filho MC: **AtPIN: Arabidopsis thaliana Protein Interaction Network.** *BMC Bioinformatics* 2009, **10**:454.

38. Cui J, Li P, Li G, Xu F, Zhao C, Li Y, Yang Z, Wang G, Yu Q, Li Y *et al.*: **AtPID: *Arabidopsis thaliana* protein interactome database — an integrative platform for plant systems biology.** *Nucleic Acids Res.* 2008, **36**:D999-1008.

39. Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH, Geisler M: **A predicted interactome for Arabidopsis.** *Plant Physiol.* 2007, **145**:317-329.

40. De Bodt S, Hollunder J, Nelissen H, Meulemeester N, Inzé D: **CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations.** *New Phytol.* 2012, **195**:707-720.

41. Van Landeghem S, Hakala K, Rönnqvist S, Salakoski T, Van de Peer Y, Ginter F: **Exploring biomolecular literature with EVEX: Connecting genes through events, homology, and indirect associations.** *Adv. Bioinformatics* 2012, **2012**:582765.

42. Müller H-M, Kenny EE, Sternberg PW: **Textpresso: an ontology-based information retrieval and extraction system for biological literature.** *PLoS Biol.* 2004, **2**:e309.

Part III: Developing an optimized AP-MS workflow in rice

Chapter 4: Picking the best building blocks for TAP in rice

A strategy to improve the state of the art of AP-MS in rice

To accomplish our goal and improve the state of the art of TAP in rice, we first defined the most important parameters that drive sensitive protein complex purifications. The most crucial elements are the promoter, which determines the expression level of the bait encoding gene, and the affinity tag, decisive for the purity and yield of the purification. We selected three different promoters. PRO_{35S} is the 35S promoter of the Cauliflower mosaic virus (CaMV 35S) [1,2]. This promoter is well established for use in dicots, but was also used in rice for TAP purposes in a study on the GIGANTEA complex [3]. PRO_{GOS2} is the promoter of the rice *GOS2* gene [4]. The GOS2 protein is 52% identical to the protein encoded by the *SUI1* gene of yeast, which is a suppressor locus of translational initiator mutants [5]. This suggests that the GOS2 gene product is part of the translation machinery. Hence it is annotated as eIF1 (for eukaryotic translation initiation factor 1) [6]. PRO_{HMG} is a *HIGH MOBILITY GROUP PROTEIN* promoter (WO2004070039 A2). In house data show that PRO_{GOS2} provides medium constitutive and ubiquitous expression. PRO_{HMG} gives weak constitutive and ubiquitous expression. The optimal promoter should enable to outcompete the non-tagged endogenous bait protein in the wild type (WT) genetic background for incorporation into protein complexes, whilst not highly exceeding the concentrations of available complex to avoid over-accumulation of uncomplexed baits.

The original TAP tag developed in yeast [7] consists of two immunoglobulin G (IgG)-binding units of protein A from *Staphylococcus aereus,* a cleavage site for the tobacco etch virus (TEV) protease and a calmodulin-binding peptide (CBP). Despite optimisation of this tag for use in plants [8], it still has some limitations. The use of the CBP domain is problematic in eukaryotic cells and purifications require EGTA-containing buffers. This results in the non-specific isolation of endogenous calmodulin-binding proteins and prevents the purification of cation-dependent enzyme complexes respectively [9]. Also the overall yield and specificity of the process is low and could be improved [9,10]. Complex purification was further optimised by testing different TAP tags in Arabidopsis cell suspension cells. Evaluation of a SFZZ TAP tag, where the CBP part in the traditional TAP tag was replaced by linear peptide epitopes, resulted in reduced background, but also in low amounts of purified complexes [9]. A major leap forward in terms of both sensitivity and yield was obtained with the implementation of the GS tag [9]. This TAP tag combines two IgG-binding domains of protein G with a streptavidin-binding peptide, separated by two TEV cleavage sites.

We designed a series of new tags and corresponding vectors based on the two-step purification principle that proved successful in TAP. Previous experiments with the GS tag [9] in rice showed significant proteolysis of the bait protein (Aurine Verkest, personal communication). Based on this observation, we opted to test a variety of TAP tags (Figure 1), utilizing combinations of the building blocks of TAP and GS tag together with new affinity domains. One group includes variants of the original TAP tag. The TAPi tag, for 'improved TAP tag', is a plant-adapted version of the traditional yeast tag, with an optimized codon sequence for plants and

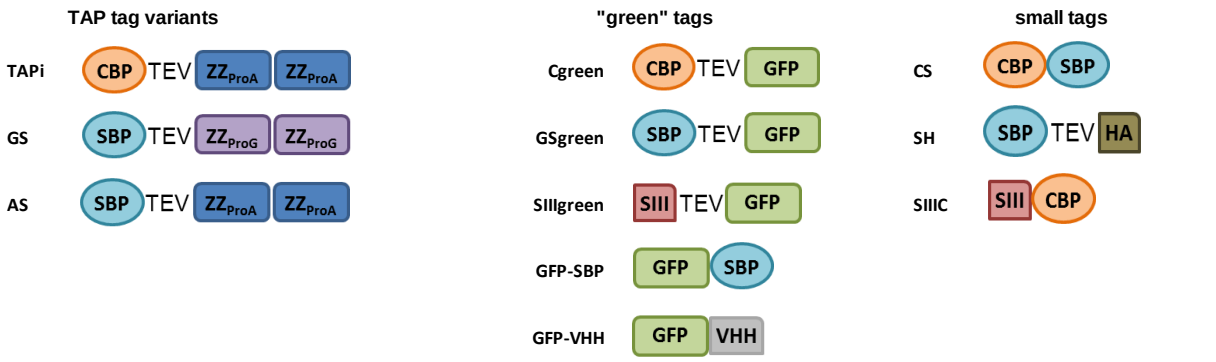
without cryptic nuclear localisation signals [8]. The AS tag is equivalent to the TAPi tag, but has the CBP domain replaced by a streptavidin-binding peptide [10], and for the GS tag the protein A portion was additionally exchanged with two IgG binding units of protein G from *Streptococcus sp.* [9,10]. We also opted to design tags containing the green fluorescent protein (GFP) and named these 'green' tags. The combination of localisation and interaction data is indeed very powerful. Using GFP as an affinity domain became feasible only recently, with the development commercially available affinity matrices having sufficient specificity for the fluorophore (www.miltenyibiotec.com, www.chromotek.com). We designed three tags analogous in design to the traditional TAP tag, but with GFP replacing the two IgG-binding domains of protein A. The first affinity domain was either the calmodulin-binding peptide, the streptavidin-binding peptide or the StrepIII peptide respectively, resulting in the Cgreen, GSgreen and SIIIgreen tags (Figure 1).

In addition, we developed the GFP-SBP and GFP-VHH tags which contain GFP and the streptavidin-binding peptide or a short amino acid sequence respectively. For the latter a VHH nanobody was developed (at the Jan Steyaert lab) that specifically recognises this epitope. This nanobody could then be used to produce suitable affinity resin. A third category of tags we considered were smaller tags. Some postulate that a smaller tag minimizes interference with the bait's function and could in that aspect enhance the success rate of the TAP protocol [11]. The commercialised CS-tag (Interplay Mammalian TAP system, Agilent technologies, Santa Clara, CA, US) consists of the calmodulin-binding peptide and streptavidin-binding peptide. The StrepIIIC tag comprises the StrepIII peptide followed by a calmodulin-binding peptide, and the SH tag contains the streptavidin-binding peptide and a hemagglutinin (HA) epitope tag with a TEV cleavage site in between. The latter yielded 30-40% purification efficiency in mammalian cell cultures [11].

We analysed the efficiency of the 3 promoters and 11 TAP tag variants by using the rice ortholog of CYCLIN-DEPENDENT KINASE REGULATORY SUBUNIT 1 (CKS1) as bait protein. Rice CKS1 is encoded by loc_os03g05300 and has 91.1% global amino acid similarity with its Arabidopsis counterpart. Arabidopsis CKS1 has a conserved role in cell cycle regulation as scaffold protein. It functions as a docking factor for positive and negative regulators of CYCLIN-DEPENDENT KINASE (CDK) activity [12–14]. These regulators were identified in a comprehensive study of the cell cycle interactome in Arabidopsis cell suspension cultures [15], some of which were confirmed in binary interaction screens [16]. Arabidopsis CKS1 associates with A- and B-type CDKs and these CDKs on their turn interact with cyclins. The core cell cycle interactome in Arabidopsis mainly represents heterodimeric interactions of CDKA1;1 with D- and A3-type cyclins, whereas B-type CDKs interact with B- and D-type cyclins [15–17]. Hence, Arabidopsis TAP experiments using CKS1 as bait resulted in co-purification of a variety of CDK/cyclin complexes. The diversity of CDK's and cyclins that co-precipitate with CKS1 can be adopted to compare the performance of the different TAP building blocks. The more CDK/cyclin complexes are co-purified using particular TAP construct, the better that construct is performing.

Figure 1. Overview of the different affinity domains and proposed combinations for TAP tag design.

domain	full name	affinity to	elution	pros	cons
CBP	calmodulin binding peptide	calmodulin	mild - EGTA addition	allows mild elution, small	EGTA required in buffers
SBP	streptavidin binding peptide	streptavidin	mild - biotin addition	allows mild elution, less background than CBP	*in vivo* biotinylated proteins can non-specifically interact with the affinity matrix
GFP	green fluorescent protein	GFP antibody	harsh - pH/boiling	provides localisation signal to the tag	bulky, requires harsh conditions for elution, expensive affinity matrix
HA	hemagglutinin	HA antibody	mild - addition of HA	allows mild elution, small	linear
ZZ_{ProG}	ZZ domain of protein G	IgG	harsh - pH/boiling	very specific binding	bulky, requires harsh conditions for elution
ZZ_{ProA}	ZZ domain of protein A	IgG	harsh - pH/boiling	very specific binding	bulky, requires harsh conditions for elution
SIII	StrepIII	Streptavidin	mild - biotin addition	allows mild elution, small	linear



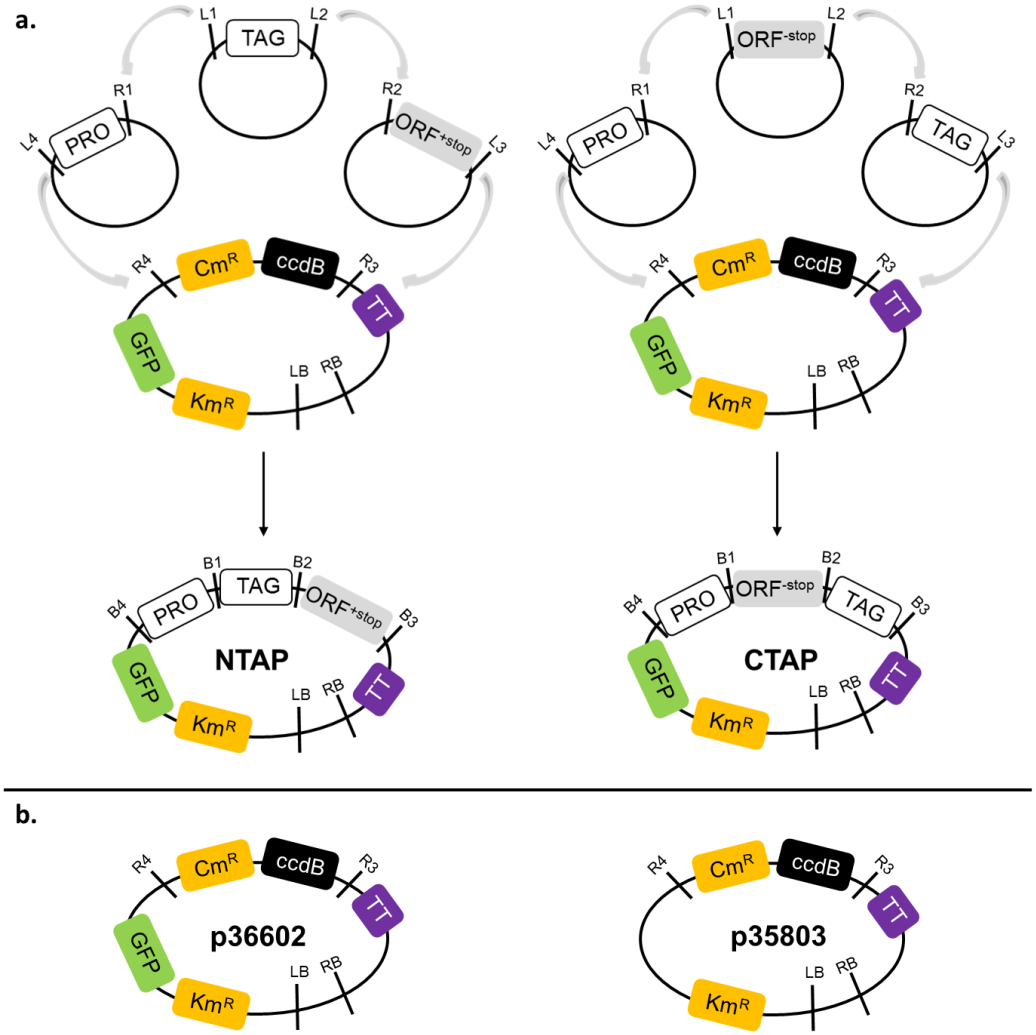
Establishment of a smooth cloning protocol

Creating a TAP construct in eukaryotic cells basically encompasses cloning of a promoter, the open reading frame (ORF) of the gene of interest and the ORF of the TAP tag into one expression vector. To maximise the chance of successfully pulling down interactors, traditionally both N- and C-terminal fusion of the TAP tag with the gene of interest are produced. We thus needed to acquire a cloning protocol that allows fast and versatile combination of the three building blocks for creating TAP cassettes.

We adopted a cloning strategy based on the MultiSite Gateway® recombination system established in Arabidopsis [18] for use in rice. To accomplish this, we cloned the Gateway® cassette from the pDEST™ R4-R3 Vector II (Life Technologies, Carlsbad, CA) into the p05050 destination vector suitable for *Oryza sativa* transformation (WO2011114279 A1). Two types of destination vectors were derived from our cloning efforts. One destination vector, p35803, contains the GFP screenable marker, whereas this cassette was removed in the p36602 vector. The latter is therefore suitable for insertion of TAP cassettes that already contain GFP as affinity domain. Our MultiSite Gateway® strategy allows cloning of the different building blocks in merely one single recombination step once promoter, gene of interest and TAP tag are present in the appropriate pENTRY vectors (Figure 2). In contrast to the strategy proposed by Van Leene *et al.* [18], our workflow requires only one type of destination vector for both N- and C-terminal tagging. Also, the

residues derived from the Gateway® cloning junction remain the same (attB2) for both fusions, ensuring consistency of the cloning products (Figure 2).

Figure 2. Overview of the MultiSite Gateway® based cloning strategy.



a. For both N- and C- terminal cloning, a three-fragment recombination strategy was used. ENTRY vectors are produced by a BP clonase-catalysed recombination step that transferred the amplicons of promoter, TAP tag coding sequence and the ORF of the gene of interest in the appropriate pDONR™ vectors. For N-terminal tagging, the coding sequence of the tag is devoid of its stop codon to allow translational fusion with the ORF. Similarly, for C-terminal tagging, the stop codon is removed from the ORF sequence. The three fragments are assembled in the pTAP destination vector in a single MultiSite LR clonase reaction to produce an expression clone. This destination vector contains within the left (LB) and right (RB) T-DNA border sequences, in addition to the Gateway® cassette, a Kanamycin resistance (Km^R) gene for selection and a GFP cassette as screenable marker. For cloning of TAP tags that contain GFP as an affinity domain, a destination vector is used that lacks the GFP screenable marker cassette. **b.** Destination vectors developed. The p36602 vector is a derivative of p35803 lacking the GFP screenable marker cassette. For clarity reasons, the Streptomycin and Spectinomycin resistance gene present in the vector backbone is not shown. TT, termination sequence from zein; Cm^R, Chloramphenicol resistance gene; ccdB, toxic killer gene for negative selection.

With an established cloning workflow in place, we created 13 TAP constructs for testing the efficiency of the different building blocks. The three different promoters were combined with a C-terminal fusion of CKS1 and the TAPi tag, cloned in the p35803 destination vector. We cloned each of the 11 TAP tag variants as C-terminal fusions with CKS1, driven by the CaMV 35S promoter. They were cloned in the appropriate destination vector according to presence (cloning in p36606) or absence (cloning in p35803) of GFP in the tag. All CKS1 test constructs had the TAP tag fused to the C-terminus, as prior knowledge from Arabidopsis showed that this translational fusion had the least effect on the CKS1 functionality [15].

Generation of rice callus cell cultured cells expressing the TAP constructs

In first instance, we used transformed rice callus cells to produce the tagged proteins. These cells enable relatively fast generation of an in principle unlimited supply of biomass. Because of their undifferentiated nature, they are an ideal biomass source to study basal pathways. Briefly, explants from dehusked rice seeds were co-cultivated with *Agrobacterium tumefaciens* strain lba4044/pal4404 harbouring binary vectors that contain a TAP fusion construct. Three days after co-cultivation, explants were separated from the seeds, washed and grown on selection medium. The selection medium contains G418 disulfate to select for transformed callus. After 2 weeks, microcalli were isolated and further proliferated on selection medium until 30g of callus was obtained. In principle, each microcallus represents an individual transformation event. We opted nonetheless to pool the different events to level out positional effects from the T-DNA insertion site [18].

Evaluation of the performance of the different TAP building blocks

The different promoters

We first analysed protein extracts from transgenic callus cell lines expressing the TAPi-tagged CKS1 constructs driven by the three different promoters by immunoblotting with α-CKS1 to confirm expression. The bait's expression level is clearly stronger when driven by the 35S CaMV compared to the other promoters (Figure 3). Also, as expected, the PRO_{GOS2} provides higher bait expression levels than PRO_{HMG} (Figure 3).

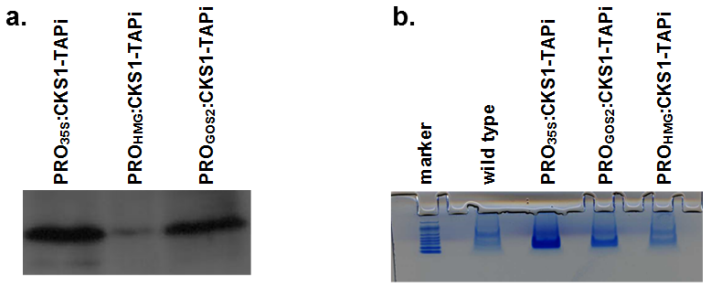
Subsequently, we performed full scale TAP experiments to gain insight on how the bait's expression levels influence the sensitivity of interactor detection. We used 50mg of total soluble protein extract for each experiment. The protocol (adapted from Rohila *et al.,* [19]) is described in detail in the materials and methods section. After purification, the samples were loaded on precast 4-12% gradient NuPAGE gels to get rid of chemical contaminants and stained with Coomassie Brilliant Blue staining. Gels were then sliced into gel plugs and subjected to *in-gel* trypsin digestion. The resulting peptides were ultimately separated on a nano-liquid chromatography (LC) column and analysed on a LTQ Orbitrap Velos mass spectrometer. The resulting mass spectra were searched against the Michigan State University (MSU) rice annotation database [6] using the SEQUEST engine. In principle we retained only proteins that were identified with at least two peptides with high confidence rank, of which at least one is unique to the protein. To

deconvolute the read-out, the resulting protein identification list was cross-checked against a preliminary list of nonspecific proteins. This preliminary background list was constructed from proteins identified in a 'mock' purification using wild type callus samples. From the resulting interaction datasets, we created a pivot table (Figure 3c). This table compiles how many CDKs and cyclins were found; interactors that typically co-purify in a CKS1 pull-down. We further appended the table with other proteins potentially related to the cell cycle. For this we added proteins with GO annotation 'cell cycle' for molecular function. *A posteriori,* we supplemented the interaction dataset with cell cycle related proteins resulting from only one significantly detected peptide.

From these three purifications, we could derive a common interaction dataset of 6 interacting proteins (Figure 3c, interactors depicted with 'x'). These include 4 CDKs, 3 of the A-type and 1 of the B-type, and 3 D-type cyclins. Participation of CKS1 in complexes involving CDKA or CDKB and CYCD is thus conserved in rice. We can also state that the TAPi tag does not interfere with CKS1 protein function. Pull-down experiments from the construct containing the CaMV 35S promoter resulted in the detection of the most significant interactors (9). These include two additional B-type CDKs that were not found in the datasets from the other two constructs. Purification using the PRO_{GOS2}-driven construct resulted in only slightly less (8) significant interactors, with 6 of the interactors overlapping with the experiment from the CaMV 35S-driven construct. Apart from the 6 common interacting proteins, we did not find additional significant identifications for purifications from the PRO_{HMG}-driven construct. This might pinpoint that bait expression is suboptimal to provide sensitive detection of interactors.

We conclude that bait expression levels in callus are sufficient for all three promoters to provide successful pull-down of cell-cycle related protein complexes. The moderate overlap between datasets can be partially explained by a lack of replicate experiments. The interaction dataset derived from the PRO_{HMG}-driven construct however shows a significant number of proteins derived from single-peptide detection (Figure 3c, interactors depicted with '1'). We therefore decided to only retain the 35S CaMV and the PRO_{GOS2} promoters for further TAP experiments in rice.

Figure 3. Comparison of the performance of three different promoters driving the expression of TAP constructs in rice callus



c.

	accession	description	symbol	PRO35S:CKS1-TAPi	PROGOS2:CKS1-TAPi	PROHMG:CKS1-TAPi	total
	loc_os03g05300	CYCLIN-DEPENDENT KINASE REGULATORY SUBUNIT 1	CKS1	8	8	4	3
cyclin-dependent kinases	loc_os03g02680	CYCLIN-DEPENDENT KINASE A 1	CDKA1	13	16	11	3
	loc_os02g03060	CYCLIN-DEPENDENT KINASE A 2	CDKA2	10	9	11	3
	loc_os01g67160	CYCLIN-DEPENDENT KINASE B 1-1	CDKB1;1	15	4	4	3
	loc_os08g40170	CYCLIN-DEPENDENT KINASE B 2-1	CDKB2;1	12			1
cyclins	loc_os05g14730	CYCLIN A 1	CYCA1;4			1	1
	loc_os12g39210	CYCLIN A 3-2	CYCA3;2			1	1
	loc_os01g59120	CYCLIN B 1-1	CYCB1;1	2			1
	loc_os04g47580	CYCLIN B 2-1	CYCB2;1		2		1
	loc_os09g21450	CYCLIN D 2-1	CYCD2;1	1	1	1	3
	loc_os09g02360	CYCLIN D 3-2	CYCD3;2	2	6	2	3
	loc_os12g39830	CYCLIN D 5-2	CYCD5;2		1	1	2
	loc_os03g10650	CYCLIN D 5-3	CYCD5;3	2	4	5	3
	loc_os07g37010	CYCLIN D 6-1	CYCD6;1	2	1		2
	loc_os11g43900	TRANSLATIONALLY-CONTROLLED TUMOR PROTEIN	TCTP	1		1	2
	loc_os03g01740	CYCLIN-DEPENDENT KINASE INHIBITOR EL2	CKI		2		1
			total sign	9	8	6	11
			total	11	11	11	16

a. Immunoblot analysis using the α-CKS1 antibody. The molecular weight of TAPi-tagged CKS1 is 32.4 kDa. **b.** NuPAGE gel stained with Coomassie Blue of tandem affinity-purified samples. **c.** Interaction dataset resulting from the TAP-MS experiments on each construct. Only known CKS1-interaction partners or proteins annotated with GO term 'cell cycle' are shown. For each protein, the number of matching (high or medium confidence) peptides retrieved in the purification are shown. Single peptide identifications would not be withheld in standard analysis, hence these are marked in grey.

The different TAP tags

Next, we tested which affinity tags suit the best to accommodate optimal TAP purifications in rice. Analysis by α-CKS1 immunoblotting on protein extracts from callus cell lines expressing the different TAP constructs displayed substantial differences in bait expression levels between the classes of tags (Figure 4). It seems that large protein-like domains tend to stabilize the bait's expression. Tags that contain the large globular IgG-binding domains showed massive expression. This was also the case for the 'green' tags, although for these affinity handles some protein decay was visible (Figure 4). The smaller tags showed quite low abundance. We anticipated this as problematic, as in our system the TAP construct

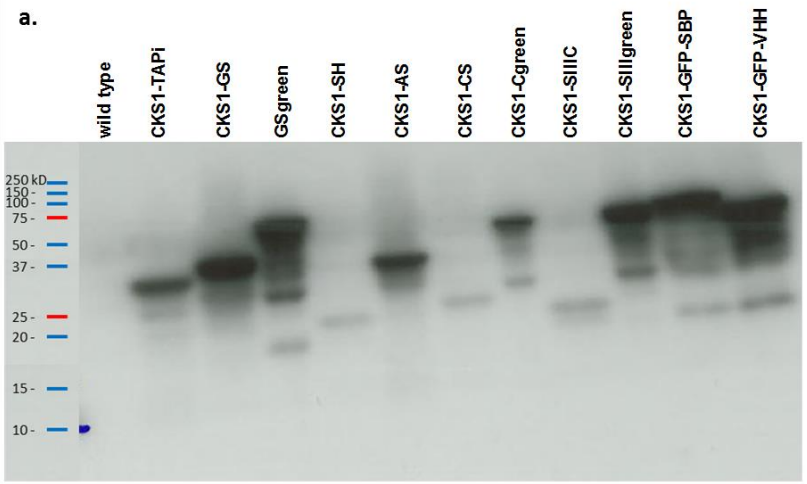
has to outcompete its endogenous counterpart for incorporation into protein complexes. We therefore decided not to retain the small tags in our toolbox.

Remarkably, in our hands the GS tag was in fact stable contrary to what was observed before (Aurine Verkest, personal communication). We hypothesize that the earlier observed instability might have been the result of a different sequence context of the Gateway® att sequences or 'bad luck' of choosing baits that are prone to degradation. From this observation, we opted to further assay the performance of two TAP tag variants that contain the SBP-domain – the GS and GSgreen tag. SBP has indeed low-nanomolar affinity to streptavidin, can be specifically eluted by (desthio)biotin, and was already shown to perform superior compared to the CBP domain in mammalian cells [10] and Arabidopsis cells [9]. The SBP domain is also preferred to the StrepIII tag as the latter is a linear tag. Linear tags can lead to interference in protein folding since they do not possess a secondary or tertiary structure. Similar to our strategy for the promoter comparison, we performed full scale TAP experiments on the GS, GSgreen and TAPi constructs. The latter was considered a comparison to the state of the art. For all purifications, 50mg of protein extract was used and purifications were performed at least in duplicate, in the most optimal conditions for each tag. For the GSgreen tag, we also explored the possibility of performing a single step purification using GFP as affinity handle. Details are provided in the materials & methods section. Sample preparation, mass spec and data analysis were performed as described for the promoter comparison. From the resulting interaction datasets, we created a pivot table summarizing how many CDKs and cyclins were found, together with proteins with GO annotation 'cell cycle' for molecular function. We included protein identifications derived from only one significantly detected peptide to also consider more weakly associated binders.

From the nine purifications, we could derive a common interaction dataset of 3 interacting proteins (Figure 4c); 2 CDKs of the A-type and 1 of the B-type. These interactors comprise the complete interaction dataset for the single-step purification using the GSgreen construct. The common dataset resulting from all two-step purifications included in addition CYCLIN D 6;1 and a cyclin-dependent kinase inhibitor. The latter was mainly identified from only one significantly detected peptide, stressing its weak interaction. We further retrieved CDKB2;1 in both TAPi and GS purifications, whereas it was only found once with one peptide matching the protein sequence in the GSgreen purifications. The interaction dataset mainly contains D-type cyclins. An explanation could be that CYCDs are simply more abundant in the purified sample as both CDKA as CDKB associate with CYCDs [17], whereas B-type cyclins only interact with CDKB [17]. Apart from the CYCLIN-DEPENDENT KINASE INHIBITOR (CKI) retrieved in all double step purifications, we found two additional CKI's. Each CKI was retrieved in only one purification, by one significantly detected peptide. Probably, this is due to a more weak or unstable nature of the interaction with the complex. Overall, two-step purifications from the TAPi-, GS- and GSgreen-constructs resulted in roughly the same amount of significant interactors (Figure 4b). Since the GS tag was shown to outperform the TAPi tag in terms of yield and sensitivity using different baits in other organisms [9,10], we decided to only retain the GS and the similar GSgreen tag for further analysis.

Figure 4. Comparison of the performance of different affinity tags in rice callus.

a.



construct	molecular weight
CKS1	10.6 kDa
CKS1-TAPi	32.4 kDa
CKS1-GS	33.4 kDa
CKS1-GSgreen	48.2 kDa
CKS1-SH	20.4 kDa
CKS1-AS	33.6 kDa
CKS1-CS	22.2 kDa
CKS1-Cgreen	44.9 kDa
CKS1-SIIIC	21.3 kDa
CKS1-SIIIgreen	45.2 kDa
CKS1-GFP-SBP	44.9 kDa
CKS1-GFP-VHH	40.2 kDa

b.

				PRO_{35S}:CKS1-TAPi			PRO_{35S}:CKS1-GS			PRO_{35S}:CKS1-GSgreen			
	accession	description	alias	repl1	repl2	repl3	repl1	repl2	repl3	repl1	repl2	ss	total
	loc_os03g05300	CYCLIN-DEPENDENT KINASE REGULATORY SUBUNIT 1	CKS1	7	8	6	8	6	9	8	9	4	9
	loc_os03g02680	CYCLIN-DEPENDENT KINASE A 1	CDKA1	9	7	6	14	14	12	14	16	7	9
	loc_os03g01850	CYCLIN-DEPENDENT KINASE A 1	CDKA1			6	13			12			3
CDKs	loc_os02g03060	CYCLIN-DEPENDENT KINASE A 2	CDKA2	6	5	4	8	7	5	9	9	4	9
	loc_os01g67160	CYCLIN-DEPENDENT KINASE B 1-1	CDKB1;1	8	2	8	16	7	8	8	10	5	9
	loc_os08g40170	CYCLIN-DEPENDENT KINASE B 2-1	CDKB2;1	1	3	8	8	4	2	1			7
	loc_os01g59120	CYCLIN B 1-1	CYCB1;1			2							1
	loc_os09g21450	CYCLIN D 2-1	CYCD2;1					1					1
cyclins	loc_os09g02360	CYCLIN D 3-2	CYCD3;2			1	4	5		1	3		5
	loc_os12g39830	CYCLIN D 5-2	CYCD5;2		1	1		2	1				4
	loc_os03g10650	CYCLIN D 5-3	CYCD5;3			1	1				2		3
	loc_os07g37010	CYCLIN D 6-1	CYCD6;1	1	2	6	2	3	2	3	2		8
	loc_os03g01740	CYCLIN-DEPENDENT KINASE INHIBITOR EL2	CKI	1	1	1	3	1	1	1	2		8
CKIs	loc_os02g52480	CYCLIN-DEPENDENT KINASE INHIBITOR 1	CKI1			1							1
	loc_os10g33310	CYCLIN-DEPENDENT KINASE INHIBITOR 4	CKI4						1				1
			total sign	4	6	8	9	8	6	6	8	4	12
			total	7	8	13	10	10	9	9	8	4	15

a. Immunoblot analysis using the α-CKS1 antibody. The molecular weights are shown in the column on the right. **b.** Interaction dataset resulting from the TAP-MS experiments on each construct. Only known CKS1-interaction partners or proteins annotated with GO term 'cell cycle' are shown. For each protein, the number of matching (high or medium confidence) peptides retrieved in the purification are shown. Single peptide identifications would not be withheld in standard analysis, hence these are marked in grey. In the last column, the total amount of times the protein was identified throughout the different purifications is shown. The exact composition of the TAP tags are shown in Figure 1. Abbreviations used: repl: replicate; ss: single step; CDKs: cyclin-dependent kinases; CKIs: cyclin-dependent kinase inhibitors.

Conclusion

Rice has become a well-established model organism for cereal crops. It is especially a most suitable model for research in the field of enhancing crop yields, being a crop itself. To date, merely individual genes regulating components of plant growth or seed yield are known, but their connections are largely unexplored.

Nevertheless, biological processes are often carried out by the dynamic interaction of proteins in complexes and signalling pathways. Unravelling these interconnections can thus greatly speed up our understanding of complex traits as yield. AP-MS is the preferred technology to capture protein complexes in planta. The technology is state-of-the-art in Arabidopsis, but not in important crop species like rice. Therefore, we optimized each of the different building blocks - promoter, TAP tag, expression vector- required for making a TAP construct.

We opted to constitutively overexpress the TAP tagged construct in rice cells, as the tagged proteins produced must be able to outcompete endogenous forms for incorporation into the protein complexes. This strategy is the most frequently used for TAP in plants and was previously shown to lead to higher complex recovery compared to expression with endogenous promoters in WT background of Arabidopsis cell suspension cultures [18] and plants [20]. Overexpression might also facilitate the detection of interaction partners with lower affinity or more transient associations. A trade-off is that overexpression might lead to aberrant interactions leading to false positive identifications. We anticipate that the constitutive nature of the promoters is not a problem when screening complexes from rice callus. These indeed represent mainly a population of undifferentiated cells. The issue could be more problematic when plants are used as biomass source, as these consist of a multitude of cell types. An alternative is to express the TAP construct under control of the endogenous promoter in null mutant background. This approach is however very costly in terms of throughput. Also, rice mutant collections are not as exhaustive as in Arabidopsis. In our case however, we want to analyse the protein complexes from proteins that show growth enhancing characteristics when overexpressed. An overexpression strategy for our TAP experiments is therefore the most logic choice.

We tested three different constitutive promoters ranging in level of overexpression, in combination with the cell cycle protein CKS1 and the TAPi tag. Immunoblot analysis of the different constructs showed the expected differences in expression level. For the weaker constitutive promoter PRO_{HMG} – the promoter of a rice *HIGH MOBILITY GROUP PROTEIN* – a significant portion of the interactors identified resulted from single-peptide identifications. These would not have been retained in standard analyses. Purification experiments from constructs containing the two other promoters, PRO_{GOS2} – the promoter of the rice *GOS2* gene- and PRO_{35S} – the Cauliflower Mosaic Virus 35S promoter – resulted in a similar amount of interaction partners and were therefore considered as equally fitting for incorporation in a TAP construct. Importantly, we only used a single experiment to assay each promoter's performance. Therefore, more elaborate screening of multiple baits in replicate experiments will be needed to further validate these preliminary results.

We next tested three groups of TAP tags, representing a total of 11 different affinity handles. For their design, we selected and combined affinity domains that showed successful in AP-MS experiments in eukaryotes. The TAP tag variants tested include the TAPi tag, shown to enable high throughput experiments in rice, the GS tag and the AS tag. The GS tag showed to outperform the TAPi tag in both final yield and specificity in mammalian and Arabidopsis cells [9,10]. Preliminary tests

in rice plants however hinted to potential instability of the tag (Aurine Verkest, personal communication), which forced us to reach out for alternatives to develop an optimised TAP protocol. An alternative group of tags have the green fluorescent protein included as affinity domain. The use of GFP as affinity domain became feasible with the recent development of affinity matrices containing efficient antibodies. A key advantage is that GFP allows both protein localization and affinity purification experiments. In that frame, a 'localisation and affinity purification' (LAP) tag was introduced in metazoans [21].

Alternatively, we appended a group of tags consisting of combinations of only small domains or linear tags to minimize the size of the affinity handle. A frequently posed concern is indeed the large size of the 21 kDa TAP tag and potential interference with complex assembly, localisation or functionality of the bait protein. In this context, a SH-tag was proposed, comprising only 48 amino acids [11]. The authors claimed to achieve overall purification yields between 30 and 40%, a significant increase compared to the 10% observed with the GS tag.

In our hands, the SH tag, and in general all small tags tested showed a significantly lower expression pattern compared to the larger affinity handles. Clearly, the presence of a large domain tends to stabilise bait accumulation. Further, comparison of the interaction profiles we retrieved from CKS1 tagged with TAPi, GS or GSgreen with the associations already described in literature hints that functionality is not compromised, although the tags were in this case double the size of the protein. A further confirmation is found from yeast data, where in 82% of the cases essential genes were replaced with a tagged version of the gene, a viable haploid strain was obtained [22]. This number is deduced from C-terminal tagging only and might this be higher if both N- and C-terminal fusions would have been tested. From these considerations we decided to drop the small tags from further testing because of their low accumulation. A further comparison of the resulting interaction datasets from two variants that contain the SBP-domain – the GS and GSgreen tag- with the TAPi tag showed similar for all three tags. Since SBP performs superior compared to the CBP domain in terms of specificity and sensitivity [9,10] we decided to only use the GS and similar GSgreen for further experiments.

Our findings should be seen as provisory, as merely one single bait was used for scoring the TAP tools. The test bait was however carefully chosen to quickly scan a broad range of tools. As scaffold protein, CKS1 interacts with multiple types of different cyclins and CDK's. We adopted the diversity of CDK's and cyclins that co-precipitate with CKS1 to compare the performance of the different TAP building blocks and pre-filter the 13 different TAP constructs to two promoters and two TAP tags. In a next phase, these tools will be further tested on a range of different baits in callus and plant tissues to further consolidate our observations.

In conclusion, we established a TAP-MS strategy for purifying protein complexes from rice callus cells and tested a set of building blocks that can contribute to an optimised protocol. We demonstrated this by identifying interactors from the cell cycle protein CKS1. These building blocks will subsequently be adapted to study the molecular mechanisms that underlie growth or seed yield in (cereal) crops, biological processes for which *Oryza sativa* is generally used as a model system.

This functional interactomics approach can help to further pave the way in gaining understanding in this complex trait.

Materials & Methods

Construction of TAP destination vectors
We used the p05050 destination vector for rice transformation (WO2011114279A1) as starting point to construct the TAP destination vectors. In first instance, the p35803 vector was created by replacing the region encompassing the GOS2 promoter and attR1-attR2 Gateway® cassette by the attR3-attR4 Gateway® cassette. The resulting p35803 vector was verified through sequencing the region between left and right border. The p36602 was subsequently derived from p35803 by removing the GFP cassette and verified through sequencing in the same manner as for p35803.

Cloning of the TAP building blocks in the appropriate pENTRY vectors
Vectors containing the PRO_{HMG}, PRO_{35S} and PRO_{GOS2} promoters were used as template for isolation of the promoter regions by high fidelity PCR using the Phusion Hot Start II high-fidelity DNA polymerase kit (Thermo Fisher Scientific, Waltham, MA). Primers were designed using VectorNTI primer design. The recombination sites for Gateway® cloning were added to the primers. To verify amplification, PCR product was loaded on a 1% agarose gel and run for 25 min at 100V. Subsequently, fragments of the corresponding size were excised from gel and extracted using the QIAquick gel extraction kit (Qiagen, Venlo, The Netherlands). Next, the promoter sequences were cloned in compatible pDONR™P4-P1R vectors (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Successful recombination was verified by restriction digest analysis and sequencing of the Gateway® cassette.

All eleven TAP tags were manufactured through gene synthesis (GeneArt®, Life Technologies, Carlsbad, MA) including the appropriate att-sites. They were cloned in compatible pDONR™P2R-P3 vectors (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Successful recombination was verified by restriction digest analysis.

The ORF for CKS1 was amplified from an expression vector that already contained the sequence. Primers were designed using the VectorNTI primer design tool with the att-sequences added, but in such a way that the CKS1 sequence would be amplified without its stop codon. PCR was performed using the Phusion Hot Start II high-fidelity DNA polymerase kit (Thermo Fisher Scientific, Waltham, MA). To verify amplification, PCR product was loaded on a 1% agarose gel and run for 25 min at 100V. Subsequently, fragments of the corresponding size were excised from gel and extracted using the QIAquick gel extraction kit (Qiagen, Venlo, The Netherlands). Next, the CKS1 amplicon was cloned in the compatible pDONR™221 vector (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Successful recombination was verified by restriction digest analysis and sequencing of the Gateway® cassette.

Generation of TAP expression vectors through MultiSite Gateway®

Recombination of the promoter, CKS1 sequence and TAP tags was performed using the standard MultiSite Gateway® cloning technology to generate the TAP destination vectors as described above. The resulting TAP expression vectors were verified by restriction digest analysis and transferred to *Agrobacterium tumefaciens* lba4044/pal4404 by electroporation. Transformed bacteria were selected on yeast extract broth medium with the appropriate antibiotics and verified by colony PCR using Takara Taq polymerase (Takara Bio Inc, Shiga, Japan) according to the manufacturer's instructions.

Transformation and callus biomass generation

Agrobacterium tumefaciens-mediated transformation of *Oryza sativa* (ecotype japonica) seeds was performed according to patent WO2001006833 A1 with minor modifications. After mechanical dehusking using rice husker Kett US TR120, 150-200 seeds were surface sterilized with 6% sodium hypochlorite solution for 45 minutes and washed with sterile water. Afterwards, seeds were transferred to callus induction medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 4 g/L gelrite, 2 mg/L 2,4-D) and allowed to germinate at 32°C under continuous light of 3000 lux. Six days after germination, the seeds were briefly submerged in liquid infection medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 68.5 g/L sucrose, 36 g/L D+ glucose-monohydrate, filter sterilised) containing 100 µM acetosyringone and transgenic A. tumefaciens lba4404/pal4404 containing the TAP destination vector (OD600 0,05-0,1) and transferred to co-cultivation medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 30 g/L sucrose, 10 g/L D+ glucose-monohydrate, 4 g/L gelrite, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 µM acetosyringone). Co-cultivation was allowed for three days at 25°C in darkness. Thereafter, the explants were removed from the seeds, washed with 250 mg/L cefotaxime and transferred to selection medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 7 g/L agarose type 1, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 35 mg/L G418 disulfate) for incubation under continuous light (3000 lux) at 32°C. Twelve days later, microcalli were isolated and transferred onto fresh selection medium, refreshed every ten days, and grown until 30 g of callus was obtained. The callus material was then harvested in liquid nitrogen and stored at -80°C for subsequent analysis.

Expression analysis of the TAP constructs

Callus material was ground to homogeneity in liquid nitrogen with mortar and pestle. About 200 µL of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM MgCl2, 150 mM NaCl, 15 mM pNitrophenyl phospate, 60 mM β- glycerophosphate, 0.1% NP-40, 0.1 mM Na3VO4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTAfree Ultra Complete tablet (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) was added and homogenized with a 1.5-mL pellet mixer. Homogenized samples were flash frozen in liquid nitrogen, thawed on ice and centrifuged twice for 15 min at 4°C at 20,800 g. Protein concentrations were determined by Bradford assay (Bio-rad, Hercules, CA). Fifty µg of total protein extract was loaded for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 0.75 mm 12% Mini-PROTEAN® TGX™ precast gels (Bio-Rad, Hercules, CA) for 20 min at 300 V in TGX running buffer (25 mM Tris-HCl, pH 8.3, 1.92M glycine, 35 mM

SDS). Resolved proteins were transferred to PVDF membranes using Trans-Blot® Turbo™ Mini PVDF transfer packs and the Trans-Blot® Turbo™ Transfer system (Bio-rad, Hercules, CA) according to instructions of the manufacturer. Blotted PVDF membranes were then incubated in blocking buffer (3% Difco™ skimmed milk (w/v) in TBS-T buffer (50 mM Tris, 150 mM NaCl pH8.0, 0.1% Triton X-100)) overnight at 4°C or 1 h at room temperature (RT) on an orbital shaker. After this blocking step, membranes were incubated for 1 h at RT with peroxidase anti-peroxidase antibody (Sigma-Aldrich, Saint-Louis, MO) in blocking buffer on an orbital shaker. Membranes were washed 1 x 15 min and 4 x 5 min with TBS-T buffer. Bound antibody was detected by mixing equal amounts of the two chemiluminescent reagents from the ECL-kit (Perkinelmer, Waltham, MA) and incubating for 1 min. Membranes were placed in a film cassette and exposed to an Amersham hyperfilm™ ECL film (GE Healthcare, Wauwatosa, WI) in a dark room, where autoradiograms were also developed.

Protein extract preparation for TAP purifications
Callus material was ground to homogeneity in liquid nitrogen with mortar and pestle. Crude protein extracts were prepared in two volumes of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM MgCl2, 150 mM NaCl, 15 mM p-nitrophenyl phosphate, 60 mM β-glycerophosphate, 0,1% NP-40, 0.1 mM Na3VO4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTA-free Ultra Complete tablet Easypack (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) at 4°C using an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC). Soluble fraction was obtained from isolating the supernatans after double centrifugation at 36,900 g for 20 min at 4°C. The extract was passed through four layers of miracloth (Merck KGaA, Darmstadt, Germany) and kept on ice.

Tandem affinity purification of TAPi-tagged bait
Extract preparation was performed according to above described, with addition of 0,5 mM EGTA to the extraction buffer.

Purifications were performed as described by Van Leene et al. (2007) [16] with some modifications. The protein extract was added to 25 µL of effective immunoglobulin G (IgG) Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 375 µL or 150 column volumes wash buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 50 µL wash buffer and 2x 10U (2 x 1 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker.

Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 100 µL calmodulin binding buffer (10 mM Tris-HCl pH8, 150 mM NaCl, 0,1% NP-40, 10 mM β-mercaptoethanol, 1 mM imidazole, 2 mM $CaCl_2$, 1 mM Mg acetate, 1 mM PMSF, 1 µM E64, 5% Ethylene glycol) was added to the beads to collect residual eluate. One mM of $CaCl_2$ was added to the

resulting eluate, and this was incubated for 1 h at 4°C under gentle rotation with 100 µL effective calmodulin agarose beads (Stratagene, La Jolla, CA), pre-equilibrated with 3x 1 mL calmodulin binding buffer. Calmodulin beads were transferred to a mobicol column and washed with 100 column volumes or 10 mL wash buffer. Complexes were eluted by adding 1 mL of calmodulin elution buffer (10 mM Tris-HCl pH8, 150 mM NaCl, 0,1% NP-40, 10mM β-mercaptoethanol, 1 mM imidazole, 25 mM EGTA, 5% Ethylene glycol). Eluate was precipitated with TCA (25% (v/v) and the resulting protein pellet was washed twice with ice-cold acetone containing 50 mM HCl.

Tandem affinity purification of GS-tagged bait.
Purifications were performed as described by Van Leene et al. (2010) [13] with minor modifications. The protein extract was added to 25 µL of effective immunoglobulin G (IgG) Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 375 µL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 50 µL wash buffer and 2x 10U (2 x 1 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 100 µL wash buffer was added to the beads to collect residual eluate. The resulting eluate was incubated for 1 h at 4°C under gentle rotation with 25 µL effective streptavidin Sepharose High Performance beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL wash buffer. Streptavidin beads were transferred to a mobicol column and washed with 100 column volumes or 2.5 mL wash buffer. Complexes were eluted in 40 µL NuPAGE sample buffer containing 20 mM desthiobiotin (Sigma-Aldrich, Saint-Louis, MO) by 5 min incubation on ice, followed by centrifugation at 1,500 rpm at 4°C.

Tandem affinity purification of GSgreen-tagged bait.
The protein extract was added to 100 µL of effective GFP-Trap® agarose beads (Chromotek GmbH, Planegg-Martinsried, Germany), pre-equilibrated with 3x 1 mL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 15 mL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 200 µL wash buffer and 2x 40U (2 x 4 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 400 µL wash buffer was added to the beads to collect residual eluate. The resulting eluate was incubated for 1 h at 4°C under

gentle rotation with 25 µL effective streptavidin Sepharose High Performance beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL wash buffer. Streptavidin beads were transferred to a mobicol column and washed with 100 column volumes or 2.5 mL wash buffer. Complexes were eluted in 40 µL NuPAGE sample buffer containing 20 mM desthiobiotin (Sigma-Aldrich, Saint-Louis, MO) by 5 min incubation on ice, followed by centrifugation at 1,500 rpm at 4°C.

Single-step affinity purification of GSgreen-tagged bait.
The protein extract was added to 100 µL of effective GFP-Trap® agarose beads (Chromotek GmbH, Planegg-Martinsried, Germany), pre-equilibrated with 3x 1 mL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 15 mL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Sample was eluted by boiling in NuPAGE sample buffer.

Sample preparation
Purified protein samples were loaded and separated with a short 7-min run on a precast 4-12% gradient NuPAGE Bis-Tris gel (Life Technologies, Carlsbad, CA), fixed in 50% EtOH/2% H3PO4 and visualized with colloidal Coomassie Brilliant Blue G-250 (Sigma-Aldrich, Saint-Louis, MO) staining.

Proteolysis and peptide isolation
NuPAGE gel containing purified protein samples was destained twice in HPLC-grade water (Thermo Fisher Scientific, Waltham, MA) for 1 h and incubated in 25 mL of reducing buffer (6.66 mM DTT plus 50 mM NH4HCO3 in HPLC-grade water) for 40 min to reduce the polypeptide disulphide bridges. Subsequently, thiol groups were alkylated by incubating the gel for 30 min in 25 mL of alkylating buffer (55 mM iodoacetamide, 50 mM NH4HCO3 in HPLC-grade water) in the dark before washing with HPLC-grade water. The zone containing the protein sample was sliced from the gel and sectioned into different gel plugs. These were washed twice with 600 µL of HPLC-grade water and dehydrated in 600 µL 95% acetonitrile twice for 10 min. The dehydrated gel plugs were submerged and rehydrated in 90 µL trypsin digest buffer (12.5 µg/mL trypsin (MS gold; Promega, Madison, WI) in 50 mM NH4HCO3 and 10% (v/v) acetonitrile in HPLC-grade water) for 30 min at 4°C. Afterwards, trypsin digestion was allowed for 3.5 h at 37°C. Resulting peptide samples were sonicated for 5 min in a sonication bath and the solution covering the gel plugs (containing trypsinized peptides) were kept aside. Remaining gel plugs were completely dehydrated in 95% acetonitrile for 10 min and the remaining acetonitrile solution was added to the first fraction of trypsin digests. The resulting trypsin-digested sample was completely dried in a SpeedVac for 2-3 h at 4°C.

LC-MS/MS analysis
A nano LC system (NanoLC Ultra 2D system, Eksigent, Dublin, CA) was connected to an LTQ Velos Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) with a trapping column (PepMap 100, C18 precolumn with 5-µm particles, 20mm × 200 µm internal diameter; Dionex), flow rate of 6 µL/min (100% Solvent

A) and 5 minutes after injection switched in line with an analytical C18 column (Acclaim PepMap 100, 3-µm particles, 150 mm × 75 µm internal diameter; Dionex). A chip-based nano-electrospray source (TriVersa, Advion Biosystems, Ithaca, NY) operated at 1.8 kV.

Peptides were solubilised in loading Solvent A (2% acetonitrile, 0.1% acetic acid (v/v) in HPLC grade water) and 10 µL of the sample was loaded on the trapping column. Peptide samples were separated with a 65 min gradient at a flow rate of 300 nL/min. MS spectra were recorded in the Orbitrap FT analyzer with a resolution of 60,000 (at m/z 400) and an automatic gain control (AGC) target setting of 500,000. The maximum injection time was set to 500 ms, and lock mass was enabled (polysiloxane ion at m/z 445.12024). Collision-induced dissociation MS/MS spectra were acquired by the ion trap in data-dependent mode, selecting up to the 20 most abundant multiply charged precursor ions from the MS spectrum. The maximum injection time was set to 50 ms and an AGC setting of 7,500. Fragmentation was accomplished by collision- induced dissociation wideband activation at normalized collision energy of 35 eV and an activation time of 30ms. After MS/MS, the m/z precursors were excluded for 30 s.

Analysis of the protein interaction data

Peak lists were generated and submitted for protein identification with Proteome Discoverer 1.3.0.339 (Thermo, Bremen, Germany). Spectrum grouping was allowed with a maximum retention time of 1 minute and a precursor mass tolerance of 2 ppm. Peak lists were generated only for MS/MS spectra containing more than 5 peaks, with a total intensity above 50. The relative signal-to-noise limit was set to 5. Peak lists were submitted for protein identification against the Michigan State University annotation database for rice containing 66,338 sequence entries with search engine SEQUEST. Enzyme was set to trypsin, allowing for maximum 1 missed cleavage site. Precursor mass tolerance was set to 3 ppm and fragment mass tolerance at 0.8 Da. Fixed modifications were set to carbamidomethylation of cysteines. Variable modifications were set to methionine oxidation and methylation of aspartic acid and glutamic acid, with a maximum of 4 modifications per peptide. Peptides were validated using a decoy database search. The strict target False Discovery Rate (FDR) was set to 0.01 (or 1%), the relaxed FDR was set to 0.05 (or 5%). Only high confident (>99%) peptides were withheld. Only proteins with at least two matched high confident peptides were retained. *A posteriori,* also proteins identified with one high-confidence peptide were added, if they were known CKS1 interactors. A list of nonspecific background proteins was assembled by combining background proteins from control purifications on mock callus extracts identified with the LC/MS setup. To obtain the final list of interactors, these background proteins were subtracted from the list of identified proteins.

Supplementary information

Figure S1. Sequence alignment of the amino acid sequence of AtCSK1 with OsCKS1.

```
ATCKS1      MGQIQYSEKY FDDTFEYRHV VLPPEVAKLL PKNRLLSENE WRAIGVQQSR
OSCKS1      MGQIQYSEKY FDDTYEYRHV VLPPEVAKLL PKNRLLSENE WRAIGVQQSR

ATCKS1      GWVHYAVHRP EPHIMLFRRP LNYQQQQENQ A---QNMLVK
OSCSK1      GWVHYAIHRP EPHIMLFRRP LNFQQQQEAA AAAAAQMLPK
```

Sequence alignment was generated using the MAFFT alignment tool [23].

Figure S2. Primers used in this chapter.

primer n°	name	primer sequence
prm24592	attB1r PRO_{HMG}	ggggacaagtttgtacaaaaaagcaggctggggactgctttttttgtacaaacttgcggctgaatcctgcgagaa
prm24591	attB4 PRO_{HMG}	ggggactgctttttttgtacaaacttgcaaatggctgaatcctgcgagaagggcg
prm24590	attB1r PRO_{GOS2}	ggggacaagtttgtacaaaaaagcaggctggggactgctttttttgtacaaacttgcgaactttgctggtgaaag
prm24589	attB4 PRO_{GOS2}	ggggacaactttgtatagaaaagttgctaatccgaaaagtttctgcaccgt
prm24588	attB1r PRO_{35S}	ggggacaagtttgtacaaaaaagcaggctggggactgctttttttgtacaaacttgcgagatagatttgtagagagagact
prm24587	attB4 PRO_{35S}	ggggacaactttgtatagaaaagttgctcgacactctcgtctactc
prm25565	attB1 CKS1 no stop	ggggacaagtttgtacaaaaaagcaggctatggccagatccagtact
prm25566	attB2 CKS1 no stop	ggggaccactttgtacaagaaagctgggtaggtacttgggcagcat

Supplementary background list and mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP) under the filenames SI_Chapter_4_background_list.xlsx, SI_Chapter_4_PRO-CKS1_MS_data and SI_Chapter_4_CKS1-TAPtag_MS_data.xlsx respectively.

Author contribution

The PhD candidate was in charge for generation of the TAP constructs, transformation, maintenance and upscaling of callus lines, expression analyses, TAP purifications, mass spectrometry data analysis and writing of the manuscript.

References

1. Franck A, Guilley H, Jonard G, Richards K, Hirth L: **Nucleotide sequence of cauliflower mosaic virus DNA.** *Cell* 1980, **21**:285–294.

2. Odell JT, Nagy F, Chua N-H: **Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter** *Nature* 1985, **313**:810–812.

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

4. De Pater BS, van der Mark F, Rueb S, Katagiri F, Chua NH, Schilperoort RA, Hensgens LA: **The promoter of the rice gene GOS2 is active in various different monocot tissues and binds rice nuclear factor ASF-1.** *Plant J.* 1992, **2**:837–844.

5. Yoon HJ, Donahue TF: **The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon.** *Mol. Cell. Biol.* 1992, **12**:248–260.

6. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al.: **Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data.** *Rice (N. Y).* 2013, **6**:4.

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

8. Rohila JS, Chen M, Cerny R, Fromm ME: **Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants.** *Plant J.* 2004, **38**:172–181.

9. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

10. Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superti-Furga G, Bauch A: **An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells.** *Nat. Methods* 2006, **3**:1013–1019.

11. Glatter T, Wepf A, Aebersold R, Gstaiger M: **An integrated workflow for charting the human interaction proteome: insights into the PP2A system.** *Mol. Syst. Biol.* 2009, **5**:237.

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

13. Boudolf V, Rombauts S, Naudts M, Inzé D, De Veylder L: **Identification of novel cyclin-dependent kinases interacting with the CKS1 protein of Arabidopsis.** *J. Exp. Bot.* 2001, **52**:1381–1382.

14. De Veylder L, Beeckman T, Beemster GT, Krols L, Terras F, Landrieu I, van der Schueren E, Maes S, Naudts M, Inzé D: **Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis.** *Plant Cell* 2001, **13**:1653–1668.

15. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, et al.: **Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana.** *Mol. Syst. Biol.* 2010, **6**:397.

16. Van Leene J, Boruc J, De Jaeger G, Russinova E, De Veylder L: **A kaleidoscopic view of the Arabidopsis core cell cycle interactome.** *Trends Plant Sci.* 2011, **16**:141–150.

17. Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E: **Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network.** *Plant Cell* 2010, **22**:1264–1280.

18. Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, et al.: **A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana.** *Mol. Cell. Proteomics* 2007, **6**:1226–1238.

19. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

20. Fàbregas N, Li N, Boeren S, Nash TE, Goshe MB, Clouse SD, de Vries S, Caño-Delgado AI: **The brassinosteroid insensitive1-like3 signalosome complex regulates Arabidopsis root development.** *Plant Cell* 2013, **25**:3377–3388.

21. Cheeseman IM, Desai A: **A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans**. *Sci STKE.* 2005, **266**:pI1.

22. Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M, et al.: **Functional organization of the**

yeast proteome by systematic analysis of protein complexes. *Nature* 2002, **415**:141–147.

23. Katoh K, Standley DM**: MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol. Biol. Evol.* 2013, **30**:772–780.

Chapter 5: Transferring an optimized TAP-toolbox for the isolation of protein complexes to a portfolio of rice tissues

Maarten Dedecker[1,2,3], Jelle Van Leene[1,2], Nancy De Winne[1,2], Dominique Eeckhout[1,2], Geert Persiau[1,2], Eveline Van De Slijke[1,2], Bernard Cannoot[1,2], Leen Vercruysse[1,2], Lies Dumoulin[3], Nathalie Wojsznis[3], Kris Gevaert[4,5] , Steven Vandenabeele[3] and Geert De Jaeger[1,2]

[1]Department of Plant Systems Biology, VIB, Technologiepark 927, 9052, Gent, Belgium; [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052, Gent, Belgium; [3]CropDesign N.V., Technologiepark 21, B-9052 Ghent, Belgium; [4]Department of Medical Protein Research and Biochemistry, VIB, Albert Baertsoenkaai 3, 9000, Ghent, Belgium; [5]Department of Biochemistry, Ghent University, Albert Baertsoenkaai 3, 9000, Ghent, Belgium

Preface

With the tools required to optimize TAP in rice selected from callus tissues, we further consolidated the technology in plant tissues. For this, we used APC10 and CDKD as bait proteins, tagged by an optimised variant of the GS tag and driven by the PRO_{35S}. We opted to employ in first instance this subset of our selected tools, as we wanted to progress in exploring plant tissues with reasonable effort. An elaborate screening of multiple baits in the callus and plant platforms using our selected tools is addressed in a next chapter (chapter 6.1).

Abstract

Proteins are the cell's functional entities that mainly interact with other proteins, rather than operating independently. Therefore, capturing *in vivo* protein complexes is crucial to gain understanding of their functioning in a cellular context. Affinity purification coupled to mass spectrometry (AP-MS) has proven to yield a wealth of information about protein complex constitutions for a broad range of organisms. For *Oryza sativa*, the technique has been initiated in callus and shoots, but has not been optimized ever since. Therefore, we translated an optimized tandem affinity purification (TAP) approach from *Arabidopsis thaliana* toward *Oryza sativa*, and demonstrated its applicability in a variety of rice tissues. A list of non-specific and false positive interactors is presented, based on re-occurrence in over more than 170 independent experiments, to cross-check bona fide interactors. We demonstrate the sensitivity of our approach by isolating the complexes for the ANAPHASE PROMOTING COMPLEX SUBUNIT 10 (APC10) and CYCLIN-DEPENDENT KINASE D (CDKD) proteins from the proliferation zone of the emerging fourth leaf. Next to APC10 and CDKD we tested several additional baits in the different rice tissues and reproducibly retrieved at least one interactor for 81.4% of the baits screened for in callus tissue and T1 seedlings.

Hence, by transferring an optimized TAP tag combined with ultrasensitive mass spectrometry, our TAP protocol enables high chances of finding interactors for a wide range of bait proteins and opens the possibility to capture complex dynamics by comparing tissues at different stages of a developing rice organ.

Introduction

Proteins are the main 'workhorse-entities' of the cells. They exert their function by participating in or affecting macromolecular assemblies, resulting in complex dynamic networks. Plant cells, that due to their sessile lifestyle need to cope with different types of environmental changes, exploit the properties of those networks to pertain homeostasis, which is translated in a huge variety of cellular processes. Understanding these processes thus requires a deep understanding of the network topology behind it. One way to gather this type of information is through the identification of protein-protein interactions (PPIs). Three methods are the main drivers for the elucidation of PPIs in plants. The yeast two-hybrid (Y2H) method identifies binary protein interactions through screening of the interaction partners in yeast. The method enables both comprehensive screening of open reading frames (ORFs), as was done for Arabidopsis [1], and a more targeted approach focusing on specific pathways or tissues [2,3]. Alternatively, one-to-one interactions can be screened within plant cells, through protein complementation analysis (PCA) [4]. Affinity purification coupled to mass spectrometry (AP-MS) identifies all proteins that co-purify with the pull-down of a tagged 'bait' protein under near-physiological conditions and thus also captures indirect interactions. The technique already proved its merits in various plant species (Arabidopsis, rice, petunia, tomato, tobacco) for different cellular processes, including the cell cycle [5], flowering [6], leaf development [7] and endocytosis [8].

Owing to its rather small (389 Mb), fully annotated genome, *Oryza sativa* is in addition to being the most important food crop in the world also an excellent model for biological research on cereals. In that frame, five AP-MS approaches have so far been presented using rice for screening PPIs. Three used cultured cells, and their performance has been proven with the isolation of interaction partners of the TATA-BOX BINDING PROTEIN (TBP) [9], GIGANTEA [10] or the *Oryza sativa* FERTILIZATION-INDEPENDENT ENDOSPERM 2-polycomb protein complex [11]. The fourth approach reported the purification of VIRESCENT YELLOW LEAF associated proteins from shoots of 6- to 8-week-old seedlings [12]. A last, more high-throughput effort was presented by Rohila and coworkers. They identified interaction partners for 23% of the 129 rice kinases screened, starting from the shoots of 6- to 8-week-old seedlings [13,14]. The low success rate in this study emphasizes that creating a comprehensive picture of possible interactions for a given complex in plants is a daunting task and that there is room for improvement of TAP protocols for plants. Indeed, plants contain a tremendous variety of cell types and cellular states, each of these shaped by specific PPI networks.

To overcome this major hurdle, we developed a more improved protocol utilizing the GS^{rhino} TAP tag. It consists of a tandem repeat of the IgG-binding ZZ domain of protein G and a streptavidin-binding peptide (SBP), separated by a tandem repeat of the specific human rhinovirus 3C (HR3C) cleavage site for gentle elution. This TAP tag is based on the GS tag that has shown a higher efficiency in terms of purification specificity and yield in higher eukaryotic cells like mammalian cells [16] or plant cells [17], as compared to the classical TAP tag developed for yeast. In Arabidopsis, combination of the GS tag and ultrasensitive MS has allowed successful optimization of the TAP procedure. In cell cultures, the approach has

enabled the identification of on average 5.6 specific interactors in common per bait protein used in a duplicate TAP experiment, with a success rate of 65% for identifying at least one interactor per bait protein, confirmed in both duplicate TAP experiments. Moreover, integration of ultrasensitive MS has allowed extrapolation of the TAP procedure to Arabidopsis seedlings to study protein complexes in a developmental context [15]. From these encouraging results in Arabidopsis, we tested the applicability of the procedure in rice, a major model for cereal crops. We applied our procedure to screen interaction partners for the ANAPHASE PROMOTING COMPLEX SUBUNIT 10 (APC10) and the CYCLIN-DEPENDENT KINASE D (CDKD) in a variety of rice tissues, including dissected organ parts.

The APC plays an important regulatory role in the eukaryotic cell cycle controlling the specificity of sister-chromatid separation and exit from mitosis by ubiquitin-mediated proteolysis of cell cycle regulators, such as CYCLIN B and SECURIN. In addition to cell cycle regulation, the APC has important roles in developmental processes in plants, as was demonstrated in Arabidopsis and rice [18–20]. Apart from two subunits required for ubiquitin ligase activity, i.e. the CULLIN-related protein APC2 and the REALLY INTERESTING NEWG GENE (RING) finger protein APC11, the APC contains at least nine additional subunits [21]. The complex is guided toward its targets by either CELL DIVISION CYCLE 20 or CCS52A (for CELL CYCLE SWITCH PROTEIN 52A) activators. These contain a WD40 protein-binding domain, which recognizes D-box, KEN-box or A-box destruction motifs [22].

CDKD was previously shown to form a heterotrimeric CDK-activating kinase (CAK) complex with a regulatory CYCLIN H (CYCH) subunit and the assembly factor 'MENAGE A TROIS 1' (MAT1) both in rice [13] and Arabidopsis [5]. In the same studies, the other sub-complex of the general transcription factor II H (TFIIH), i.e. the five-subunit core consisting of XERODERMA PIGMENTOSUM B (XPB), p34, p52, p62 and p44, was co-purified, together with another helicase subunit XERODERMA PIGMENTOSUM D (XPD), which links both complexes of the TFIIH. As part of the CAK complex, CDKD not only phosphorylates the C-terminal domain of RNA polymerase II, but also the T-loop of CDKs [23,24]. CDKD itself is thought to be activated by a CAK-activating kinase, CDKF [25].

With our protocol, we screened the APC10- and CDKD-containing complexes in different tissues and developmental contexts, i.e. callus, regenerated shoot, seedling and proliferative tissue. We retrieved the core complexes throughout the different tissues and believe that the proposed analytical procedure will set a benchmark for assaying protein complex constitutions, from cultured cells to different developmental contexts in crop plants. The proposed workflow allows for screening of multiple baits in a portfolio of different rice tissues and provides a success rate of up to 81.4% of the baits screened for.

Results

Developing a portfolio of TAP-MS workflows to study rice protein complexes

Targeted screening of protein complexes through AP-MS is based on four main steps: cloning, generation of sufficient biomass producing the tagged bait protein, protein complex purification and identification of the co-purified proteins through

mass spectrometry. We developed a rice AP-MS workflow by optimizing each step and streamlining them into one efficient process. On top, we examined a portfolio of plant tissues for their efficiency to express the bait, enabling to screen complexes in the most suitable cellular environment, depending on the biological question and on prior knowledge of the bait protein. Our platform is built from the following consecutive steps: (i) flexible and Gateway®-compatible cloning, (ii) versatile generation of plant material producing the bait protein, (iii) performant affinity purification to increase complex recovery and protocol sensitivity, (iv) liquid chromatography coupled to tandem mass spectrometry, and (v) data analysis for identification of purified complex components. In the coming sections, we will describe all individual steps of the workflow and document its performance for systematic protein complex analysis, as shown for the APC10 and CDKD complexes and covering different plant tissues.

Construction of TAP-fusion cassettes

Traditionally, a TAP construct consists of a desired promoter driving the expression of a translational fusion of the affinity tag and the protein of interest. Since the affinity tag can interfere with the function of the bait protein, both N- and C-terminal fusions are tested. This increases the chance to obtain a protein fusion that maintains functionality, improving the success rate of the purifications. We constructed a rice-specific destination vector – named pTAP – compatible with MultiSite Gateway® recombination-based cloning for both N- and C-terminal tagging (Figure 1a). This destination vector contains in between the left and right T-DNA border sequences a kanamycin resistance gene (Km^R) for selection and the Gateway® cassette followed by the termination sequence from zein. A selected promoter sequence and the coding sequences for either bait or affinity tag are first cloned into the appropriate entry vectors and subsequently recombined into the destination vector as shown in Figure 1a.

Based on earlier experiments in *Arabidopsis thaliana* [5], we made only an N-terminal fusion for APC10 and a C-terminal fusion for CDKD. The TAP tag we employed was the recently developed GS^{rhino} tag that already has proven its superiority in Arabidopsis [15]. The tag. We opted to use the Cauliflower Mozaic Virus 35S promoter to drive expression of our fusion construct. This in contrast to previous studies in rice, which used the ubiquitin promoter derived from maize [9,13,14]. We argued that the lower activity of the 35S promoter in monocot tissues (10x less in maize, Christensen *et al.* [26] would prevent overaccumulation of non-complexed bait, ensuring a higher sensitivity of the protocol. Moreover, being constitutive, 35S guarantees bait expression in all tissues. Although overexpression is simple and favours competition with the endogenous counterpart for incorporation in the complex, some concerns might be raised regarding the possibility to induce protein misfolding, mislocalisation and/or misregulation at the cellular level. However, when comparing the interaction profiles we retrieved with the associations already described in literature, this hardly seems to be the case.

Establishing a portfolio of rice tissues

We assayed the interaction partners for our two baits in four different tissues, varying throughout rice plant development (Figure 1c). In all cases, we used a series of different transformation events to level out positional effects from the T-

DNA insertion site on bait expression, as previously suggested [27]. For fast generation of sufficient cells expressing the tagged bait, we developed a protocol based on the generation of transgenic callus tissue (Figure 1c). The required amount of callus tissue is grown within 3.5 months after co-cultivation. Since callus cells are kept undifferentiated through the hormone balance in the medium, they mainly support basal pathways during their lifecycle. To detect interaction partners that show a more discrete expression pattern during development, proteins should be extracted from differentiated plant tissues rather than from undifferentiated cultured cells. We initially tested shoots immediately regenerated from callus and grown for two weeks after regeneration ('T0 shoots') (Figure 1c). Acquiring sufficient biomass using T0 shoots takes nearly as long as generating sufficient callus tissue. The trade-off is a high dependence on the regeneration efficiency. Therefore, we tested the use of 2-week-old T1-seedlings (Figure 1c). Since we went over one generation to grow seedlings, segregation of the offspring resulted in a mixed population containing homozygous, heterozygous and null individuals. To enrich for transgenic individuals, we grew the plants on medium containing the selective agent, analogous to growing cultured cells and T0 shoots. However, by sampling a whole seedling, different tissues, each containing several cell types, will be mixed. Since these cells vary in relevance to the bait's function and the presence of its interaction partners, extraction from whole plants will lead to a dilution of relevant extract, or to false positive identifications when proteins that normally do not occur together, but possess affinity for each other, might be identified as interactors. To increase the specificity of the protocol, we therefore explored the purification of complexes from the proliferating zone of the emerging fourth leaf (Figure 1c). Of course, in principle any type of tissue could be assayed, as long as enough complex can be purified to allow detection through MS.

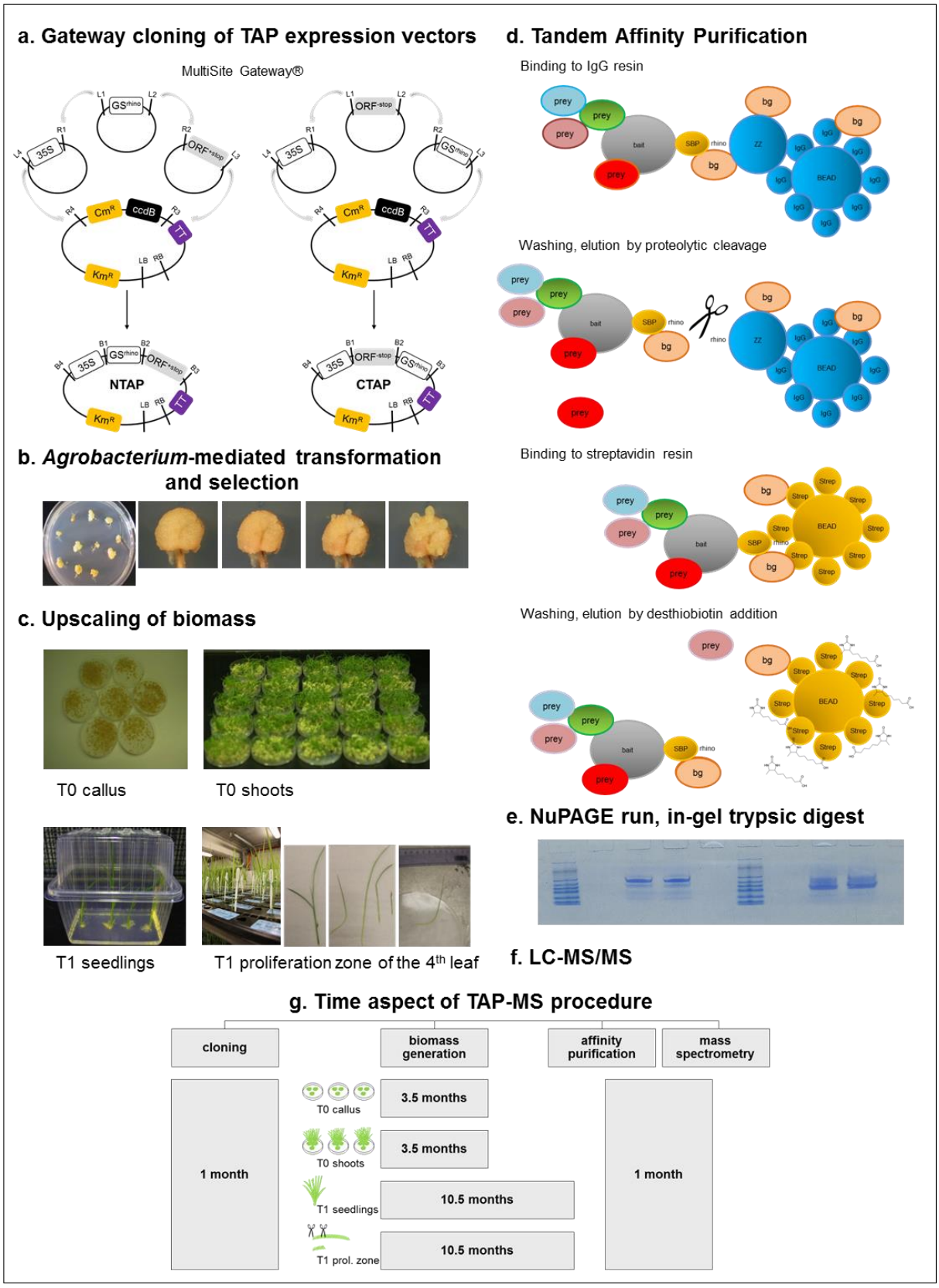
Purification of complexes from rice tissues

Prior to purification, we confirmed bait accumulation by western blotting in all tissues tested (Figure S1). From every sample type, except for T0 shoots, sufficient protein extract was generated to perform two successful independent purifications in parallel. Protein input per experiment ranged from 7.5 mg protein extract from the proliferation zone over 50 mg from callus tissues to 150 mg from seedlings. The protein input from shoots varied depending on regeneration efficiency (130 mg protein extract for APC10 and 50 mg for CDKD). Because interactors found in two repeat experiments are more reliable as compared with interactors found only once, the technical repeats help minimizing false positive identifications caused by sample handling. Protein complexes were isolated through a recently improved version of the GS tag, the GS^{rhrino} tag [15] (Figure 1d). The HR3C protease is low-temperature active [28], being a significant advantage compared to the TEV protease used with the original GS tag that requires 16°C for sufficient activity, since incubation for the proteolysis step at 4°C provides higher chances to preserve more unstable interaction partners. Bait and associated prey proteins are first retained by IgG-sepharose beads through binding with the ZZ domain of the tagged protein, followed by release from the immuno-precipitated ZZ domain through addition of HR3C protease and enrichment for a second time using streptavidin sepharose beads. Finally, they are eluted by addition of desthiobiotin due to competitive binding. We added desthiobiotin immediately to the sample buffer. Consequently, the eluate can be applied directly on gel for further sample

preparation MS. We performed in-gel trypsin digestion (Figure 1e) prior to highly sensitive MS analysis (Figure 1f). For callus and proliferation zone samples, the first purification step was performed 'in-batch', meaning that the affinity beads were simply added to the extract for binding. For shoots and seedlings, the relatively larger extract (around 25 mL on average) was brought consecutive times onto a column containing the affinity resin of the first affinity binding step, in order to optimize binding with the bait protein in the large extraction volume.

Figure 1. Overview of the TAP procedure.

Figure 1. Overview of the TAP procedure



a. Schematic representation of the strategy for cloning TAP constructs using MultiSite Gateway®. Both for N- and C-terminal cloning, a three fragment recombination strategy is performed, requiring only one type of destination vector that is suitable for any of the fusions. TT, zein-terminator; 35S, Cauliflower Mozaic Virus 35S promoter; Km^R, kanamycin resistance gene for selection of transformed callus; LB and RB, resp. left and right border for T-DNA insertion; CcdB, toxic killer gene for negative selection; Cm^R, chloramphenicol resistance gene. **b.** Illustration of the formation of microcalli after *Agrobacterium*-mediated gene transfer. Microcalli are isolated and transferred to selective medium. **c.** Overview of the different types of biomass sources tested. T0 callus tissue and T0 shoots regenerated from callus are harvested immediately after co-cultivation and sufficient growth. For T1 seedlings and T1 proliferative tissue from the fourth leaf, regenerated plants were fully grown for seed harvesting. These seeds were then used for biomass generation. **d.** Overview of the TAP purification strategy as modified from Van Leene *et al.* (2015). In short, extracted proteins are incubated with IgG resin. Binding through the IgG-binding ZZ domain of the bait retains both bait and its associated proteins. Next, retained proteins are washed to remove contaminating background (bg) proteins and gently eluted by specific cleavage through the addition of human rhinovirus 3C (HR3C) protease (scissors). The eluate is then incubated with streptavidin resin binding the remaining affinity handle, i.e. the streptavidin-binding peptide (SBP). The captured protein complexes are washed to remove residual HR3C protease and contaminating proteins, and finally eluted through competitive binding with desthiobiotin. **e.** Residual chemicals that could interfere with the MS analysis are removed by a short run on a NuPAGE gel. Afterwards, the gel slice containing the proteins is cut out and incubated with trypsin. **f.** Peptides derived from the gel slices are separated on a nano-LC column prior to analysis on the mass spectrometer. **g.** Schematic representation of the time aspect of the different steps of the TAP-MS protocol starting from different plant resources.

Establishment of a subtraction list of non-specific and false positive binders: separating the wheat from the chaff for co-purified interactors in rice

The detection sensitivity of true interactors with the bait protein during AP-MS experiments is typically reduced by contaminant proteins which bind non-specifically to the beads or the tag. In our TAP approach, these are kept at minimal levels by applying the two-step purification strategy, in which non-specific binders to the ZZ domain and the IgG-sepharose matrix that remain after washing, are physically separated from the bait and its associated proteins during the first proteolytic cleavage elution step, followed by a second washing and elution step through competitive binding with desthiobiotin.

To filter remaining non-specific bait binders that complicate interpretation of the results, we built a dataset of re-occurring proteins generated from a large set of various baits analyzed with the same purification protocol. Background contaminants that bind to the tag and/or beads or non-specific interactors of bait proteins such as household proteins (e.g. chaperones, ribosomal proteins, cytoskeletal proteins, protein translation factors, etc.) are consistent across purifications. We exploited this characteristic by compiling all interaction data from in total 174 TAP experiments with a multitude of baits. To minimize potential bias resulting from baits expected to function in related pathways or biological processes, we created 34 'bait classes' and considered proteins found with more than two different classes of baits as non-specific or background binders. A similar approach was recently successfully applied in Arabidopsis [15]. The majority of the purifications considered were performed on callus tissue (115 purifications), while purifications from plant tissue varied for shoots, seedlings and proliferation zone

with respectively 25, 26 and 8 experiments. As this dataset represent a smaller fraction of experiments derived from plants, we assigned a separate cut-off for determining non-specific proteins in experiments performed on plant tissues. Here, the cut-off was set for proteins present in more than one different class of baits. Importantly, this way of filtering discriminates specific from non-specific interactors rather than *bona fide* from background identifications. This assumes we will miss proteins that are genuinely in common between seemingly unrelated processes. True positive interactors that are present in the re-occurrence background list could be further retained by adding (semi-) quantitative data [28].

The resulting list of background and non-specific interactors that we removed from our interactor lists in rice holds up to 951 potential contaminants (Table S1), allowing more efficient filtering of non-specific proteins (Figure S2) as compared to a previous background list of 152 proteins [14]. The remaining identified proteins were considered specific for a bait and from these specific interactors, in principle only experimentally confirmed interactions were retained (Figure 2).

Figure 2. Proteins identified with our TAP protocol for CDKD and APC10 baits.

	accession	description	symbol	callus (2)	T0 shoots (1)	T1 seedlings (2)	proliferation zone (2)	Rohila 2006 (1)
CDK-activating kinase complex	loc_os05g32600	CYCLIN-DEPENDENT KINASE D 1;3	CDKD1;3	2	1	2	2	x
	loc_os03g52750	CYCLIN H-1	CYCH1	2	1	2	2	x
	loc_os11g28350	MENAGE A TROIS 1	MAT1	2	1	2	2	x
	loc_os06g07480	MENAGE A TROIS 1	MAT1	2	1	2	1	
TFIIH sub-complex	loc_os05g05260	XERODERMA PIGMENTOSUM D	XPD	2	1	2	2	x
	loc_os02g03340	TFIIH p34 subunit		2	1	2	2	x
	loc_os04g42990	TFIIH p44 subunit		2	1	2	2	x
	loc_os04g58350	TFIIH p52 subunit		2	1	2	2	x
	loc_os08g25060	TFIIH p62 subunit		2	1	2	2	x
	loc_os06g22820	CDKF-1/CAK1AT	CDKF1	1				

	accession	description	symbol	callus (2)	T0 shoots (1)	T1 seedlings (2)	proliferation zone (2)	destruction box
	loc_os05g50360	ANAPHASE PROMOTING COMPLEX 10	APC10	2	1	2	2	
TPR-lobe	loc_os06g41750	ANAPHASE-PROMOTING COMPLEX 3	APC3	2	1	2	2	
	loc_os03g13370	ANAPHASE PROMOTING COMPLEX 6	APC6	2		1	2	
	loc_os05g05720	ANAPHASE-PROMOTING COMPLEX 7	APC7	2	1	2	2	
	loc_os02g43920	ANAPHASE-PROMOTING COMPLEX 8	APC8	2	1	2	2	D-box (RxxLxxxxN)
lid	os05g0354300	ANAPHASE PROMOTING COMPLEX 1	APC1	2	1	2	2	
	loc_os02g54490	ANAPHASE PROMOTING COMPLEX 4	APC4	2	1		2	
	loc_os12g43120	ANAPHASE PROMOTING COMPLEX 5	APC5	2	1	2	2	D-box (RxxLxxxxN)
catalytic core	loc_os04g40830	ANAPHASE-PROMOTING COMPLEX 2	APC2	2	1	2	2	
	loc_os03g19059	ANAPHASE-PROMOTING COMPLEX 11	APC11	2				
regulatory subunits	loc_os03g03150	CELL CYCLE SWITCH PROTEIN 52A 1	CCS52A1	1		2		C-box (DRFIP)
	os10g0575950	SAMBA		2				
	loc_os02g10920	zinc finger family protein				2	1	
	loc_os06g07090	AP-1 complex subunit gamma-1			1	1		
	loc_os03g61160	expressed protein		2				
	loc_os03g24220	VILLIN 2	VLN2			2		GxEN-box
	loc_os03g18130	asparagine synthetase				2		
	loc_os09g36300	LON PROTEASE 2	LON2			2		
	loc_os07g07490	YT521-B domain protein					2	KEN-box, GxEN-box
	loc_os03g06240	YT521-B domain protein					2	
	loc_os11g10060	SEUSS	SEU				2	
	loc_os06g07210	RIBONUCLEOTIDE REDUCTASE 1	RNR1				2	

Detected baits (in blue) and proteins co-purifying with the baits (in black) are shown. Accession numbers are from the MSU database, except for SAMBA and APC1, which are from the RAP database. In principle, co-purifying proteins are only shown when confirmed in more than one experimental repeat (in black). CDKF is also shown however (in grey), since there is additional evidence from the SPRK-motif present in the protein sequence and a previously reported interaction from Y2H analysis [25]. The amount of times the protein was identified for each tissue is shown, with the number of replicate experiments for each tissue shown between brackets. For CDKD, the interaction data is compared with data from rice shoots [13]. From this study, proteins were identified from only one significant peptide. For APC10, the presence of a potential APC recognition motif in the interactor's protein sequence is specified.

Investigation of the complex composition throughout different tissues

To benchmark our method, we followed the constitution of two already established complexes throughout plant development using the four proposed types of plant tissues. The CAK complex of the TFIIH containing CDKD was unravelled both in rice shoots and in Arabidopsis cell suspensions by TAP [5,13]. The composition of the conserved APC in plants has been identified by TAP experiments with Arabidopsis cells suspension cultures and seedlings [5,29]. In the latter experiment, the plant-specific APC regulator SAMBA was used as bait protein instead of APC10. We performed a total of seven purifications for each bait protein, covering two technical TAP repeats on extracts derived from callus, seedlings and tissues from the proliferation zone, and a single experiment on extracts from T0

shoots. The obtained mass spectra from the Q Exactive mass spectrometer were applied for searches using the Mascot search engine against both the rice annotation project (RAP) database [30] and against the Michigan State University (MSU) rice database [31], containing 82,162 and 66,338 entries, respectively. Only proteins identified with at least two significant peptides, of which one is unique, were retained, and non-specific and background proteins were filtered out as described above. Figure 2 shows co-purified proteins that were confirmed in at least two independent TAP purifications. For both complexes, we were able to retrieve the interaction partners identified from previous experiments, with on top novel candidate interaction partners, indicating the higher sensitivity of our protocol.

CDKD

In first instance, we confirmed the participation of CDKD into the general TFIIH complex. TFIIH is responsible for two separate functions in eukaryotes: it melts the DNA around a lesion during nucleotide excision repair and helps to open the DNA template during the process of gene transcription. The complex consists of two sub-complexes: a trimeric sub-complex containing CDKD, CYCH and MAT1, and a core complex built from XPB, p34, p52, p44 and p62; both sub-complexes are linked by the helicase XPD. Similar to previous reports from rice leaves [13], we retrieved only eight of the nine expected subunits of TFIIH. We could not retrieve the XPB subunit. Three dimensional structure studies of yeast and mammalian TFIIH show that XPB is the most distal to the CAK trimer, and only linked to the complex through binding with p52. Probably this association is too weak to withstand the lengthy TAP protocol.

The complex composition was clearly very stable throughout plant development, since we were able to retrieve all components in all assayed tissues, including the leaf proliferation zone. In addition, we identified an alternative MAT1 assembly factor as potential interactor and purified the CDKF1 activating subunit. The interaction with the latter represents a kinase-substrate interaction [32], which is a typical transient short-living interaction, explaining why we found it only once in a callus purification experiment. In rice, the direct interaction of CDKD with CDKF1 has been detected earlier using a Y2H assay [25].

APC

The APC is a highly conserved multi-subunit E3 ligase complex required for sister chromatin separation during anaphase and establishment of the G1 phase in the cell cycle [24]. A TAP-MS study using cell suspension cultured cells identified at least 11 APC subunits in Arabidopsis [5]. Three dimensional reconstruction of human APC has shown that the complex adopts a triangular shape, a bit similar-looking as an open shell [33]. The backbone contains a lobe of tetratricopeptide (TPR) domain-containing subunits APC3, APC6, APC7 and APC8, and a platform build out of APC1, APC4 and APC5. The APC10 subunit we used as bait protein functions in recognizing and recruiting D-box containing proteins for ubiquitination [34]. APC10 docks on the APC through interaction with the tetratricopeptide domain of APC3. In Arabidopsis thaliana, APC3 is encoded by two isoforms, namely APC3a/CDC27a and APC3b/HOBBIT, whereas rice contains only one APC3

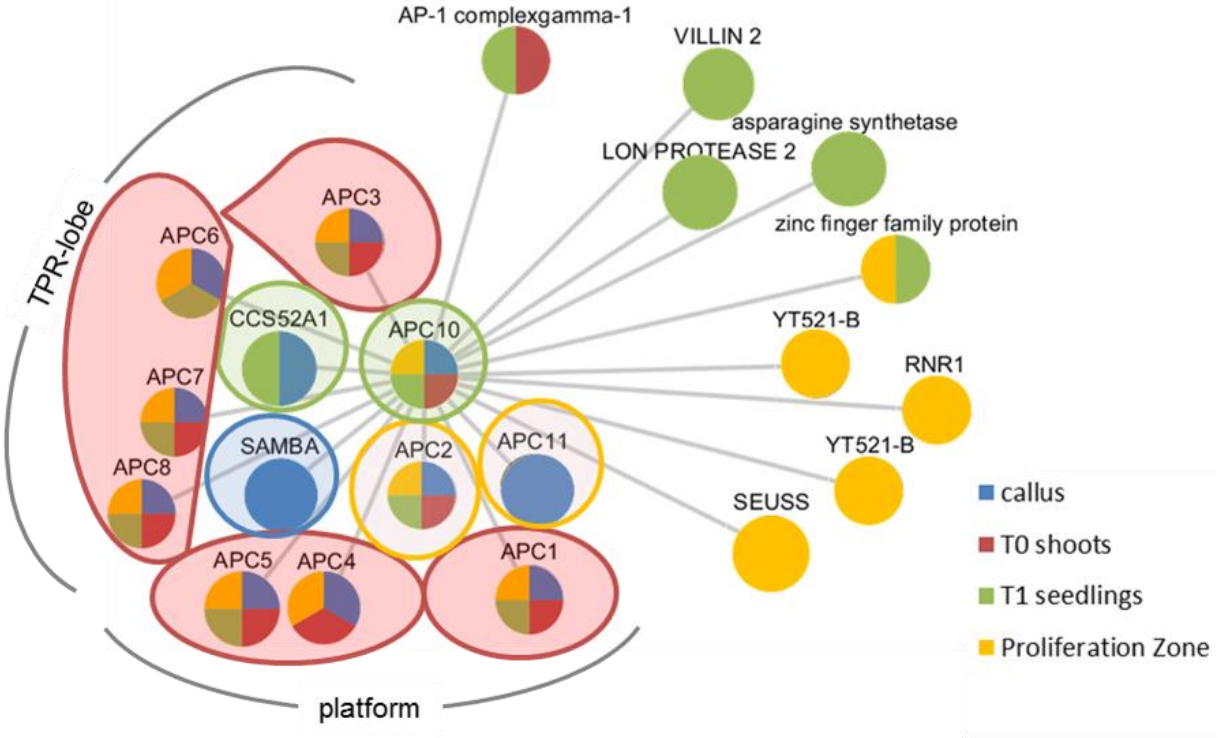
ortholog. The catalytic core of APC is built from APC2, a CULLIN domain subunit, and APC11, a RING domain subunit.

Our experiments largely confirmed this complex constitution in rice (Figure 2). In callus, containing a population of mainly dividing cells, we purified all complex constituents found in Arabidopsis, together with the recently identified APC regulator called SAMBA [29]. We detected the APC11 subunit, previously reported to be notoriously difficult to identify from TAP purifications because of its small size [29]. Also the known APC activator CCS52A was identified in callus tissue and seedlings, however only once in callus. This can be due to the very cell cycle-specific and temporal expression pattern of the protein [34].

The retrieval of the APC from complicated plant tissues (shoots and seedlings) was more challenging even with the use of more than double the amount of protein extract (130 and 150 mg, respectively) as compared to callus tissue (50 mg), since sampling of whole plants results in a mixed population of different cell types, hampering the detection sensitivity of interactors. We failed for example to detect the APC11 subunit from both shoots and seedlings, APC6 from shoots, and APC4 from seedlings (Figure 2). Also none of the CCS52 activators were retrieved from shoots, but CCS52A1 was identified and experimentally validated from seedlings.

By using specifically the proliferation zone of the leaf instead of whole T0 seedlings or T1 shoots, the ratio of relevant proliferating tissue is favoured, leading to the purification of the whole core of the APC (Figure 3, red highlighted subunits), even with a six times lower protein input as compared to callus tissues. We were also able to purify two new intriguing potential interaction partners (Figure 2). SEUSS is a well-known transcriptional co-regulator involved in flower development [35,36]. However, it was recently also found to play a more general role in lateral organ patterning [37]. In addition, we found two YT521-B domain-containing proteins, one of which contains the GxEN-box destruction motif [38].

Figure 3. Visualization of the detected APC10 interactors over different rice tissues.



Nodes are coloured according to the tissues in which they were detected. The APC holo-complex as discovered in this study is highlighted and coloured according to their function: subunits highlighted in red are APC/C backbone subunits, catalytic subunits are marked orange, activator subunits in green and negative regulators in blue. TPR-lobe: tetratricopeptide-lobe.

Discussion

Rice is, next to Arabidopsis, a very successful model species in plant biology. Quite large sets of genetic, molecular and genomic resources are already available, and being a monocot and a crop species itself, it also provides an excellent model for cereal biology. One of the key aspects in understanding biological processes is identifying the interactions between proteins to form complexes. Different technologies have been developed to screen for these types of interactions, such as Y2H, PCA and AP-MS. The former two certainly have shown their merits in the identification of binary interactions; AP-MS is complementary to Y2H and PCA, since it isolates and identifies protein complexes rather than binary interactions.

Several studies reported the development of an AP-MS approach in rice, two based on purification from cultured cells [9,10], and three based on experiments on 6- to 8-week-old seedlings [12–14]. However, the majority only purified one complex, which makes it difficult to evaluate these individual platforms. Only the two studies on rice kinases [13,14] demonstrated the screening of a significant amount of baits, with 23% of these baits yielding actual interaction partners. Their low success rate can however partly be attributed to the screening of kinases, which might be a tricky class of proteins to screen interactors from, and the fact that they only tested N-terminal fusions. Both parameters might have

compromised the true potential of their platform. Rohila *et al.* used the traditional TAP tag developed in yeast to design their bait proteins. It was stipulated however that the calmodulin-binding peptide (CBP) affinity domain in that tag might cause troubles when used in higher eukaryotic cells. The requirement of calmodulin and EGTA-containing buffers indeed can result respectively in the non-specific isolation of endogenous calmodulin-binding proteins and prevents the purification of cation-dependent enzyme complexes [17]. In that context, the GS tag has been developed, in which the CBP domain was replaced by a streptavidin-binding domain. In addition, the ZZ domain of protein A was replaced with that of protein G, resulting in a TAP tag that outperforms the traditional TAP tag in terms of specificity and complex yield in higher eukaryotes including plants [16,17]. In a later stage, the TEV protease cleavage site in the GS tag was replaced with the HR3C cleavage site for improved protein complex stability during purification [15]. We implemented these recent technical advances and combined them with the latest and most sensitive MS technology to develop an AP-MS protocol in rice that significantly improves the state of the art of AP-MS in rice. This is reflected in the significantly higher success rate we obtained. Next to APC10 and CDKD, several additional baits were tested in the different rice tissues, showing a success rate of 81.4 % of the baits screened for in callus tissue and T1 seedlings.

These improvements are required to satisfy needs not only for screening complexes in tissues that provide sufficient protein extract, such as cultured cells, but also for screening in more technically demanding tissues, such as whole plants, or even specifically isolated organs or tissues. To our opinion, this versatility is key for elucidating biological processes, since protein complexes are known to be dynamic rather than static entities. The APC, for example, is constitutively present over the plant's life cycle, but has been shown to exist in different constitutions during development in Arabidopsis [22]. Dependent on the developmental context, different interaction partners will result in a different biological signalling and outcome. As a proof of concept, we implemented our AP-MS protocol in rice callus tissues, shoots immediately regenerated from the callus, 2-week-old seedlings and the proliferation zone of the emerging fourth leaf, and show the constitution of the APC and the CDKD-associated complex in these different tissues. We postulate that all tissues tested have their merit. Cultured cells in the form of callus tissue provide fast means for an unlimited supply of protein extract. Since they represent only one (undifferentiated) cell type, their protein extract will be equally relevant for obtaining complexes. This reasonably low sample complexity, together with the absence of functional chloroplasts (and in particular RubisCo), results in not too much sensitivity requirements from the mass spectrometer, which is reflected in the identification of the smaller SAMBA and APC11 proteins as interaction partners of APC10 and of the more transient interactor CDKF for the CDKD-associated complex.

Since undifferentiated cells mainly support active division and basal pathways, interaction partners that are expressed more discretely in developmental context might be missed. To circumvent this, interaction data from callus tissue can be complemented with interaction data from plant samples. In first instance, we tested two types of 'brute force' biomass sources, containing young areal plant tissues. Shoots immediately regenerated from callus (T0 shoots) provide a way to

have this more complex tissue type in a comparable timeframe as callus material. The downside is that we suffered from a high dependency on the regeneration efficiency for the generation of sufficient shoot biomass. Nevertheless, we were able to retrieve all core interactors with CDKD and the majority of the core subunits from the APC. To circumvent this dependency, we generated seeds from 60 individual transgenic events and selected lines having only one transgene copy for growing of plants. TAP experiments with 2-week-old seedlings revealed the core CDKD complex and, apart from APC11 and APC4, also the APC's backbone and catalytic sub-complexes. We co-purified some additional potential APC-interactors. One of these, VILLIN 2, contains the GxEN destruction box and could therefore be a potential substrate for APC-mediated proteolysis. VILLINs are a class of actin-bundling proteins that can cross-link adjacent actin filaments into bundles [39,40], which serve as preferred tracks for myosin-dependent movement of organelles and provide structure to the cytoplasm [41]. Since the APC is active in dividing or endoreduplicating cells requiring dynamic actin organisation, VILLIN 2 as an APC-target clearly makes sense. Another, LON2 is a multifunctional ATP-dependent protease which exists in bacteria, archaea and within organelles in eukaryotic cells [42]. In Arabidopsis, LON2 was found involved in peroxisome biogenesis and maintenance of function. The protease function of LON2 is required for degradation of unnecessary proteins, whereas chaperone function is involved in normal folding or assembly of proteins that may contribute to proper action of peroxisomes [43]. Remarkably, Arabidopsis LON2 showed to have a mitotic-specific activation motif and a cell cycle related expression profile, hinting towards a genuine interaction with APC [44].

A combination of lowering the sample complexity and isolating interaction partners from the relevant developmental and anatomical context is provided when the specific tissue of interest is isolated and used for purification. We provided a proof of concept by isolating the proliferation zone of the emerging fourth leaf for performing our TAP experiments. Both the complexes containing CDKD and APC10 were isolated with a rather small input of 7.5 mg of protein, extracted from the proliferation zone of the fourth leaf of 550 seedlings. Especially for the APC10, we detected interesting potential interaction partners (Figure 2). The two YT521-B-domain-containing proteins could link APC function to selective removal of specific mRNAs. The homologs in Arabidopsis are the ECT-proteins (for EVOLUTIONARY CONSERVED C-TERMINUS). The conserved C-terminal domain in these proteins is in fact the YT521-B domain, and was found to be required and sufficient for interaction with CALCINEURIN B-LIKE-INTERACTING PROTEIN KINASE 1 (CIPK1), and translocation to the nucleus [45]. This domain, the YTH-domain, is conserved across all eukaryotes. YT521-B was found to modulate alternative splice-site selection in a concentration-dependent way in human cells [46]. Further, the YT521-B homology domain present in Mmi (for MEIOTIC MRNA INTERCEPTION) was shown to be responsible for the selective removal of meiosis-specific transcripts during vegetative growth in fission yeast [47]. It could be possible that in plants, the APC regulates plant YT521-B concentrations to regulate gene expression, or even to control the expression of specific splice variants. SEUSS, another potential APC10-interaction partner, is primarily known as a transcriptional co-regulator involved in flower development [35], but was also found to act in leaf organ patterning [37]. The APC was previously shown to control

the stability of Arabidopsis transcription factors [48]. Although speculative, this illustrates the power of our protocol in uncovering new links that will help to understand and elucidate biological processes *in planta*.

In conclusion, we transferred an improved the AP-MS procedure to rice and exploited these advances to provide a portfolio of possible biological tissues to screen interaction partners for proteins of interest. This updated TAP workflow enables a success rate of retrieving interactors for a wide range of bait proteins of 81.4%. The portfolio of rice tissues that can be assayed range from tissues that provide a large amount of protein extract such as cultured cells, to more technically demanding tissues, such as whole plants and even specifically isolated parts of the plant. The latter should be envisioned as a proof of concept for assaying protein complexes from any plant organ. Indeed, our optimized protocol in combination with the ultrasensitive MS now allows to identify complexes from minute samples. This opens possibilities for elucidating biological processes by comparing protein complexes assayed from different organs or from organs at different developmental stages. This could be key in gaining a comprehensive view on the biology behind the interactors, since protein complexes are known to be dynamic rather than static entities. For this, rice is a more suitable model compared to Arabidopsis, since assaying plant organs in Arabidopsis is more difficult, since some of the plant's organs are too small at the stages when e.g. proliferation is occurring.

Materials & Methods

Construction of TAP expression vectors

The destination vector for creating TAP expression constructs was derived from p05050, a destination vector used for *Oryza sativa* transformation (WO2011114279A1) by replacing the attR1-attR2, the GFP cassette and the GOS2 promoter region by the attR3-attR4 Gateway® cassette. The final destination vector pTAP was verified by sequence analysis.

Sequences for the N-terminal and C-terminal GS^{rhino} tag with the necessary recombination sites (att sites) for Gateway® cloning were created by gene synthesis. The Gateway® cassettes containing the N-terminal and C-terminal GS^{rhino} tag were cloned in compatible pDONR™221 and pDONR™P2R-P3 vectors (Life Technologies, Carlsbad, CA) respectively according to the manufacturer's instructions.

ORFs from genes of interest were isolated with Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA) on complement DNA (cDNA) of *O. sativa* (ecotype japonica). Primers used to isolate the different ORFs were designed using vector NTI software (Life Technologies, Carlsbad, CA) and sequences of the necessary recombination sites (att sites) for Gateway® cloning were added to the primers. To verify amplification, PCR products were loaded on a 1% agarose gel and run for 25 min at 100V. Subsequently, fragments of the corresponding gene size were excised from gel and then extracted using the QIAquick gel extraction kit (Qiagen, Venlo, The Netherlands). Next, ORFs were cloned in compatible pDONR™221 (for C-terminal tagging) or pDONR™P2R-P3 (for N-terminal tagging) vectors (Life Technologies, Carlsbad, CA) according to the

manufacturer's instructions. Further cloning was performed using the standard MultiSite Gateway® cloning technology to generate the TAP destination vectors. Entry vectors were verified by sequence analysis. TAP expression vectors were then transferred to *Agrobacterium tumefaciens* lba4404/pal4404 by electroporation. Transformed bacteria were selected on yeast extract broth medium with the appropriate antibiotics and verified by colony PCR with Takara Taq polymerase (Takara Bio Inc, Shiga, Japan) according to the manufacturer's instructions.

Generation of plant tissues expressing the TAP constructs

A. tumefaciens-mediated transformation of *O. sativa* seeds was done according to patent WO2001006844 A1 with minor modifications. After mechanical dehusking using rice husker Kett US TR120, 150-200 seeds were surface sterilized with 6% sodium hypochlorite solution for 45 minutes and washed with sterile water. Afterwards, seeds were transferred to induction medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 4 g/L gelrite, 2 mg/L 2,4-D) and allowed to germinate at 32°C under continuous light of 3000 lux. Six days after germination, the seeds were briefly submerged in liquid infection medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 68.5 g/L sucrose, 36 g/L D+ glucose-monohydrate, filter sterilised) containing 100 µM acetosyringone and transgenic *A. tumefaciens* lba4404/pal4404 containing the TAP destination vector (OD_{600} 0,05-0,1) and transferred to co-cultivation medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 30 g/L sucrose, 10 g/L D+ glucose-monohydrate, 4 g/L gelrite, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 µM acetosyringone). Co-cultivation was allowed for three days at 25°C in darkness. Thereafter, the explants were removed from the seeds, washed with 250 mg/L cefotaxime and transferred to selection medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 7 g/L agarose type 1, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 35 mg/L G418 disulfate) for incubation under continuous light (3000 lux) at 32°C. Twelve days later, microcalli were isolated and transferred onto fresh selection medium, refreshed every ten days, and grown until 30 g of callus was obtained. The callus material was then harvested in liquid nitrogen and stored at -80°C for subsequent analysis.

For the generation of T0 shoot material, the same transformation protocol was utilized as described above. After isolation, the microcalli were transferred onto pre-regeneration medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 500 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 7 g/L agarose type 1, 2 mg/L kinetin, 1 mg/L α-naphthalene acetic acid, 5 mg/L abscissic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 20 mg/L G418 disulfate) and incubated for one week at 32°C under continuous light (3000 lux). Resistant callus was first brought to regeneration medium I (pH 5.8, 4.3 g/L MS salts, 1 mL/L MS vitamins, 2 g/L CasaminoAcids, 30 g/L sucrose, 30 g/L sorbitol, 10 g/L agarose type 1, 2 mg/L kinetin, 0.02 mg/L α-naphthalene acetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 20 mg/L G418 disulfate) for seven days and then transferred to regeneration medium II (pH 5.8, 4.3 g/L MS salts, 1 mL/L MS vitamins, 2 g/L CasaminoAcids, 30 g/L sucrose, 30 g/L sorbitol, 7 g/L agarose type 1, 2 mg/L

kinetin, 0.02 mg/L α-naphthalene acetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 20 mg/L G418 disulfate) for two additional weeks at 32°C under continuous light (lux 3000). Plants whose shoot and root grew more than 1 cm in length were ultimately transferred to growth medium (pH 5.8, 2.15 g/L MS medium micro and macro, 0.5 mL (0.5 x) B5 vitamins, 10 g/L sucrose, 0.05 mg/L α-naphthalene acetic acid, 0.75 g/L $MgCl_2.6H_2O$, 2.5 g/L gelrite) and incubated for two weeks at 32°C under continuous light (3000 lux) before being harvested in liquid nitrogen and stored at -80°C for subsequent analysis.

Transgenic seeds containing the TAP expression vector were derived from the transformation protocol generating shoots described above. Instead of harvesting, 60 plants, each derived from an individual transformation event, were transferred to the greenhouse and further grown until seeds could be harvested. For growing seedlings, seeds were first dehusked and sterilized as previously described before sowing them on growth medium containing the selective agent (pH 5.8, 2.15 g/L MS medium micro and macro, 0.5 mL (0.5 x) B5 vitamins, 10 g/L sucrose, 0.05 mg/L α-naphthalene acetic acid, 0.75 g/L $MgCl_2.6H_2O$, 7 g/L agarose, 20 mg/L G418 disulfate). T1 seedlings were grown in a growth chamber under short day conditions at 32°C, and harvested two weeks after sowing in liquid nitrogen and stored at -80°C for subsequent analysis. Plants for isolation of the proliferation zone were grown in jiffies in the greenhouse under short day light conditions. Once the 4^{th} leaf started emerging (approximately after two weeks), leafs were carefully separated and the 4^{th} leaf was collected. The first cm from the base was then collected in liquid nitrogen.

Expression analysis of the bait proteins

Plant material was ground to homogeneity in liquid nitrogen with mortar and pestle. About 200 µL of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM MgCl2, 150 mM NaCl, 15 mM pNitrophenyl phospate, 60 mM β- glycerophosphate, 0.1% NP-40, 0.1 mM Na3VO4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTA-free Ultra Complete tablet (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) was added and homogenized with a 1.5-mL pellet mixer. Homogenized samples were flash frozen in liquid nitrogen, thawed on ice and centrifuged twice for 15 min at 4°C at 20,800 g. Protein concentrations were determined by Bradford assay (Bio-rad, Hercules, CA). Fifty µg of total protein extract was loaded for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 0.75 mm 12% Mini-PROTEAN® TGX™ precast gels (Bio-Rad, Hercules, CA) for 20 min at 300 V in TGX running buffer (25 mM Tris-HCl, pH 8.3, 1.92M glycine, 35 mM SDS). Resolved proteins were transferred to PVDF membranes using Trans-Blot® Turbo™ Mini PVDF transfer packs and the Trans-Blot® Turbo™ Transfer system (Bio-rad, Hercules, CA) according to instructions of the manufacturer. Blotted PVDF membranes were then incubated in blocking buffer (3% Difco™ skimmed milk (w/v) in TBS-T buffer (50 mM Tris, 150 mM NaCl pH8.0, 0.1% Triton X-100)) overnight at 4°C or 1 h at room temperature (RT) on an orbital shaker. After this blocking step, membranes were incubated for 1 h at RT with peroxidase anti-peroxidase antibody (Sigma-Aldrich, Saint-Louis, MO) in blocking buffer on an orbital shaker. Membranes were washed 1 x 15 min and 4 x 5 min with TBS-T buffer. Bound antibody was detected by mixing equal amounts of the two chemiluminescent reagents from the ECL-kit (Perkinelmer, Waltham, MA) and

incubating for 1 min. Membranes were placed in a film cassette and exposed to an Amersham hyperfilm™ ECL film (GE Healthcare, Wauwatosa, WI) in a dark room, where autoradiograms were also developed.

Tandem affinity purification of protein complexes

Callus material was ground to homogeneity in liquid nitrogen with mortar and pestle. Plant material (shoots, seedlings and tissues from the proliferation zone) was ground to homogeneity in liquid nitrogen with a hand blender (Braun GmbH, Kronberg, Germany). Crude protein extracts were prepared in two volumes of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM $MgCl_2$, 150 mM NaCl, 15 mM *p*-nitrophenyl phosphate, 60 mM β-glycerophosphate, 0,1% NP-40, 0.1 mM Na_3VO_4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTA-free Ultra Complete tablet Easypack (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) at 4°C using an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC). Soluble fraction was obtained from isolating the supernatans after double centrifugation at 36,900 g for 20 min at 4°C. The extract was passed through four layers of miracloth (Merck KGaA, Darmstadt, Germany) and kept on ice.

Purifications were performed as described by Van Leene *et al.* (2015) [15] with minor modifications. For material from callus material and tissues from the proliferation zone, all purifications were performed in batch. The protein extract was added to 25 µL of effective immunoglobulin G (IgG) Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 mL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 15 mL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 100 µL wash buffer and 2x 10U (2 x 1 µL, second boost after 30 min) rhinovirus 3C protease (GE Healthcare, Wauwatosa, WI) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2-mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 100 µL wash buffer was added to the beads to collect residual eluate. The resulting eluate was incubated for 1 h at 4°C under gentle rotation with 25 µL effective streptavidin Sepharose High Performance beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL wash buffer. Streptavidin beads were transferred to a mobicol column and washed with 100 column volumes or 2.5 mL wash buffer. Complexes were eluted in 40 µL NuPAGE sample buffer containing 20 mM desthiobiotin (Sigma-Aldrich, Saint-Louis, MO) by 5 min incubation on ice, followed by centrifugation at 1,500 rpm at 4°C.

For seedling and shoot material, the first affinity purification step was performed on a Poly-Prep column (Bio-Rad, Hercules, CA) containing 100 µL of effective IgG Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 1 mL extraction buffer, with a peristaltic pump (GE Healthcare, Wauwatosa, WI), at flow rate 1 mL/min. The Poly-Prep column was then mounted to a two-way valve in a vacuum manifold system for washing with 150 mL or 15

column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 200 µL wash buffer and 2x 40U (2 x 4 µL, second boost after 30 min) rhinovirus 3C protease (GE Healthcare, Wauwatosa, WI) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 400 µL wash buffer was added to the beads to collect residual eluate. The rest of the protocol proceeded similarly as described above.

Purified protein samples were loaded and separated with a short 7-min run on a precast 4-12% gradient NuPAGE Bis-Tris gel (Life Technologies, Carlsbad, CA), fixed in 50% EtOH/2% H3PO4 and visualized with colloidal Coomassie Brilliant Blue G-250 (Sigma-Aldrich, Saint-Louis, MO) staining.

Proteolysis and peptide isolation

NuPAGE gel containing purified protein samples was destained twice in HPLC-grade water (Thermo Fisher Scientific, Waltham, MA) for 1 h and incubated in 25 mL of reducing buffer (6.66 mM DTT plus 50 mM NH_4HCO_3 in HPLC-grade water) for 40 min to reduce the polypeptide disulphide bridges. Subsequently, thiol groups were alkylated by incubating the gel for 30 min in 25 mL of alkylating buffer (55 mM iodoacetamide, 50 mM NH_4HCO_3 in HPLC-grade water) in the dark before washing with HPLC-grade water. The zone containing the protein sample was sliced from the gel and sectioned into different gel plugs. These were washed twice with 600 µL of HPLC-grade water and dehydrated in 600 µL 95% acetonitrile twice for 10 min. The dehydrated gel plugs were submerged and rehydrated in 90 µL trypsin digest buffer (12.5 µg/mL trypsin (MS gold; Promega, Madison, WI) in 50 mM NH4HCO3 and 10% (v/v) acetonitrile in HPLC-grade water) for 30 min at 4°C. Afterwards, trypsin digestion was allowed for 3.5 h at 37°C. Resulting peptide samples were sonicated for 5 min in a sonication bath and the solution covering the gel plugs (containing trypsinized peptides) were kept aside. Remaining gel plugs were completely dehydrated in 95% acetonitrile for 10 min and the remaining acetonitrile solution was added to the first fraction of trypsin digests. The resulting trypsin-digested sample was completely dried in a SpeedVac for 2-3 h at 4°C.

LC-MS/MS analysis

A nano LC system (Ultimate 3000 RSLC nano, Dionex, Amsterdam, The Netherlands) was set up connected in-line to the Q-Exative (Thermo Fisher Scientific, Waltham, MA), with a trapping column (PepMap 100) of 100-µm internal diameter (I.D.) x 20 mm (length) with 5-µm C18 Reprosil-HD beads (Dionex, Amsterdam, The Netherlands), a flow rate 6 µL/min and, 5 min after injection, switched in-line with a reverse phase analytical separating column (Acclaim, PepMap 100) of 75 µm I.D. x 150 mm with 3 µm C18 Reprosil-HD beads (Dionex, Amsterdam, The Netherlands).

Peptides were solubilized in 15 µL of loading Solvent A (0.1% (v/v) trifluoroacetic acid, 2% (v/v) acetonitrile in HPLC grade water), and 5 µL of the peptide sample was loaded on the trapping column. Peptide samples were

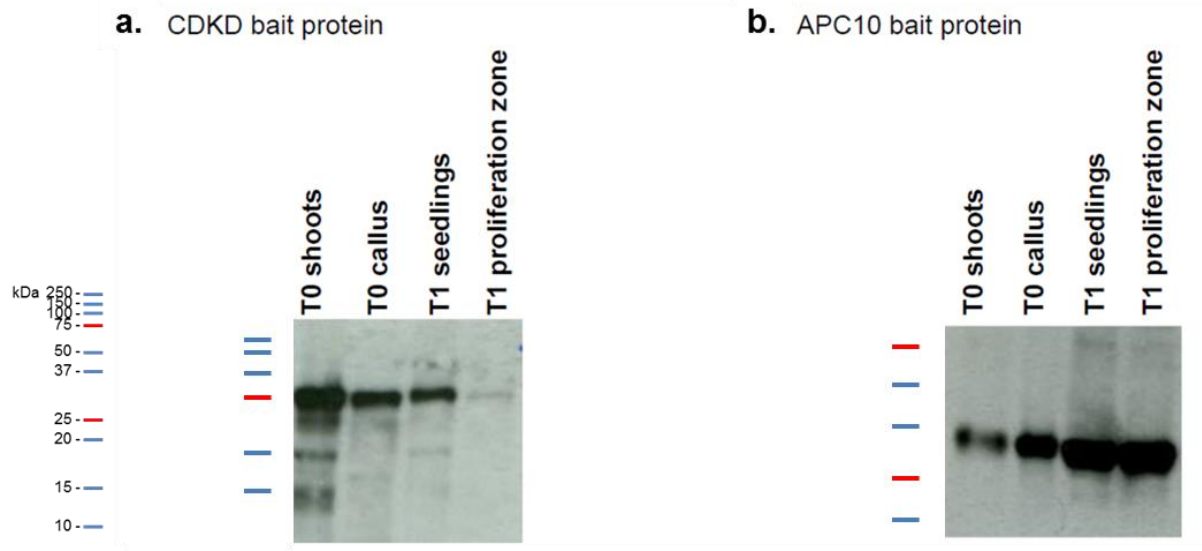
separated with a 30-min gradient from 2% mobile phase solvent A' (0.1% (v/v) formic acid in HPLC-grade water) to 50% mobile phase solvent B' (0.1% (v/v) formic acid and 80% acetonitrile in HPLC-grade water) at a flow rate of 300 nL/min. The Q Exactive was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition for the ten most abundant peaks in a given MS spectrum. Full-scan MS spectra were acquired in the Orbitrap at a target value of 1E6 with a resolution of 60.000. The ten most intense ions were isolated for fragmentation in the quadrupole, with a dynamic exclusion of 20 seconds. Target value for filling the quadrupole was set to 1E4 ion counts.

Analysis of the protein interaction data

Mascot generic files were created with the Mascot Distiller software. Grouping of spectra with a maximum intermediate retention time of 30 s and 0.005-Da precursor tolerance was allowed together with a maximum intermediate scan count of five if possible. No de-isotoping was used and the relative signal-to-noise limit was set to 2. A peak list was generated only when the MS/MS spectrum contained more than ten peaks. The Mascot Daemon interface was used to search peak lists with the Mascot search engine against the rice annotation project (RAP) database [30] and against the Michigan State University rice database (MSU) [31]. Search parameters in Mascot Daemon were as follows: variable modifications set to methionine oxidation and methylation of aspartic acid and glutamic acid, fixed modifications to carbamidomethylation of cysteines, mass tolerance on MS 10 ppm, MS/MS tolerance 20 mmu, ESI-QUAD as instrument, 2+ and 3+ as peptide charge and protease trypsin/P, allowing for 1 missed cleavage. In the Mascot result URL, a Mascot select summary was created with following settings: significance threshold $p>0.01$, maximum number of hits AUTO, Mudpit scoring, Ion score or expectancy cut-off >0.01, require bold red. Identifications were retained when at least two peptides matched a high confidence rank, with at least one peptide unique to the protein. The resulting protein identification list was cross-checked against the list of non-specific proteins. Only specific binders were retained in the final list.

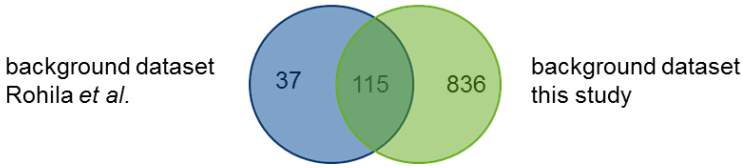
Supplementary information

Figure S1. Expression analysis of the bait proteins in the different rice tissues.



Protein extracts of the transgenic lines were analyzed by immunoblotting with peroxidase anti-peroxidase antibody against the GS^{rhino} tag to verify protein levels of the CDKD (**a.**) and APC10 (**b.**) bait. Molecular marker values are provided in the left panel in kilo Dalton (kDa). The molecular weight of the tagged proteins is 69.8 kDa for CDKD and 43.5 kDa for APC10.

Figure S2. Overlap between background datasets.



The overlap of potential background binders from this study compared to the background dataset reported in Rohila *et al.* [14] is shown.

Supplementary background list, interaction data and mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP). The established background list based on re-occurrence of a protein over independent experiments from different bait groups can be found under filename SI_Chapter_5_background_list.xlsx. The table of APC10 and CDKD-interactors, including non-confirmed interaction partners can be found in SI_Chapter_5_Full_Interaction_Datasets.xlsx. The mass spec files for APC10 and CDKD interaction data described in this chapter can be found in SI_Chapter_5_APC10_MS_data.xlsx and SI_Chapter_5_CDKD_MS_data.xlsx respectively.

Acknowledgements

Maarten Dedecker is supported by a predoctoral fellowship from the Agency for Innovation by Science and Technology and CropDesign N.V. ('Baekeland' funding, IWT 100222). The authors thank Annick Bleys for help in preparing the manuscript.

Author contribution

The PhD candidate generated the TAP constructs, did the transformations, maintained and upscaled callus cell lines and plant material, analysed the mass spectrometry data. TAP purifications and mass spectrometry measurements were mainly done by employees of VIB. PhD candidate wrote the manuscript, JVL and GDJ commented on the manuscript.

References

1. Braun P, Carvunis AR, Charloteaux B, Dreze M, Ecker JR, Hill DE, Roth FP, Vidal M, Galli M, Balumuri P, et al.: **Evidence for network evolution in an Arabidopsis interactome map.** *Science* 2011. **333:** 601-607.

2. Seo Y-S, Chern M, Bartley LE, Han M, Jung K-H, Lee I, Walia H, Richter T, Xu X, Cao P, et al.: **Towards establishment of a rice stress response interactome.** *PLoS Genet.* 2011, **7**:e1002020.

3. Lumba S, Toh S, Handfield L-F, Swan M, Liu R, Youn J-Y, Cutler SR, Subramaniam R, Provart N, Moses A, et al.: **A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in Arabidopsis.** *Dev. Cell* 2014, **29**:360–372.

4. Boruc J, Van den Daele H, Hollunder J, Rombauts S, Mylle E, Hilson P, Inzé D, De Veylder L, Russinova E: **Functional modules in the Arabidopsis core cell cycle binary protein-protein interaction network.** *Plant Cell* 2010, **22**:1264–1280.

5. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, et al.: **Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana.** *Mol. Syst. Biol.* 2010, **6**:397.

6. Smaczniak C, Immink RGH, Muiño JM, Blanvillain R, Busscher M, Busscher-Lange J, Dinh QDP, Liu S, Westphal AH, Boeren S, et al.: **Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:1560–1565.

7. Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han S-K, Jégu T, Archacki R, Van Leene J, Andriankaja M, et al.: **ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development.** *Plant Cell* 2014, **26**:210–229.

8. Gadeyne A, Sánchez-Rodríguez C, Vanneste S, Di Rubbo S, Zauber H, Vanneste K, Van Leene J, De Winne N, Eeckhout D, Persiau G, et al.: **The TPLATE adaptor complex drives clathrin-mediated endocytosis in plants.** *Cell* 2014, **156**:691–704.

9. Zhong J, Haynes PA, Zhang S, Yang X, Andon NL, Eckert D, Yates JR, Wang X, Budworth P: **Development of a system for the study of protein-protein interactions in planta: characterization of a TATA-box binding protein complex in Oryza sativa.** *Journal of Proteome Research.* **2:** 514-522.

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

11. Nallamilli BRR, Zhang J, Mujahid H, Malone BM, Bridges SM, Peng Z: **Polycomb group gene OsFIE2 regulates rice (Oryza sativa) seed development and grain filling via a mechanism distinct from Arabidopsis.** *PLoS Genet.* 2013, **9**:e1003322.

12. Dong H, Fei G-L, Wu C-Y, Wu F-Q, Sun Y-Y, Chen M-J, Ren Y-L, Zhou K-N, Cheng Z-J, Wang J-L, et al.: **A rice virescent-yellow leaf mutant reveals new insights into the role and assembly of plastid caseinolytic protease in higher plants.** *Plant Physiol.* 2013, **162**:1867–1880.

13. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

14. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

15. Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, et al.: **An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes.** *Nat. Protoc.* 2015, **10**:169–187.

16. Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superti-Furga G, Bauch A: **An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells.** *Nat. Methods* 2006, **3**:1013–1019.

17. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

18. Vanstraelen M, Baloban M, Da O, Cultrone A, Lammens T, Brown SC, Veylder L De, Mergaert P, Kondorosi E: **APC/C^{CCS52A} complexes control meristem maintenance in the Arabidopsis root.** *Proc. Natl. Acad. Sci. U. S. A.* **106:** 11806-11811.

19. Xu C, Wang Y, Yu Y, Duan J, Liao Z, Xiong G, Meng X, Liu G, Qian Q, Li J: **Degradation of MONOCULM 1 by APC/C(TAD1) regulates rice tillering.** *Nat. Commun.* 2012, **3**:750.

20. Lin Q, Wang D, Dong H, Gu S, Cheng Z, Gong J, Qin R, Jiang L, Li G, Wang JL, et al.: **Rice APC/C(TE) controls tillering by mediating the degradation of MONOCULM 1.** *Nat. Commun.* 2012, **3**:752.

21. Lima MDF, Eloy NB, Pegoraro C, Sagit R, Rojas C, Bretz T, Vargas L, Elofsson A, de Oliveira AC, Hemerly AS, et al.: **Genomic evolution and complexity of the Anaphase-promoting Complex (APC) in land plants.** *BMC Plant Biol.* 2010, **10**:254.

22. Eloy NB, Coppens F, Beemster GTS, Hemerly AS, Ferreira PCG: **The Arabidopsis anaphase promoting complex (APC): regulation through subunit availability in plant tissues.** *Cell Cycle* 2006, **5**:1957–1965.

23. Yamaguchi M, Umeda M, Uchimiya H: **A rice homolog of Cdk7/MO15 phosphorylates both cyclin-dependent protein kinases and the carboxy-terminal domain of RNA polymerase II.** *Plant J.* 1998, **16**:613–619.

24. Fabian-Marwedel T, Umeda M, Sauter M: **The Rice Cyclin-Dependent Kinase–Activating Kinase R2 Regulates S-Phase Progression** *Plant Cell.* 2002, **14**:197-210.

25. Ding X, Richter T, Chen M, Fujii H, Seo YS, Xie M, Zheng X, Kanrar S, Stevenson RA, Dardick C, et al.: **A rice kinase-protein interaction map.** *Plant Physiol.* 2009, **149**:1478–1492.

26. Christensen AH, Sharrock RA, Quail PH: **Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation.** *Plant Mol. Biol.* 1992, **18**:675–689.

27. Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N, et al.: **A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana.** *Mol. Cell. Proteomics* 2007, **6**:1226–1238.

28. Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, Dinesh-Kumar SP, Deng XW: **An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation.** *Plant J.* 2005, **41**:767–778.

29. Eloy NB, Gonzalez N, Van Leene J, Maleux K, Vanhaeren H, De Milde L, Dhondt S, Vercruysse L, Witters E, Mercier R, et al.: **SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:13853–13858.

30. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang C, Iwamoto M, Abe T, et al.: **Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics.** *Plant Cell Physiol.* 2013, **54**:e6.

31. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al.: **Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data.** *Rice (N. Y).* 2013, **6**:4.

32. Shimotohno A, Umeda-Hara C, Bisova K, Uchimiya H, Umeda M: **The plant-specific kinase CDKF;1 is involved in activating phosphorylation of**

cyclin-dependent kinase-activating kinases in Arabidopsis. *Plant Cell* 2004, **16**:2954–2966.

33. Chang L, Zhang Z, Yang J, McLaughlin SH, Barford D: **Molecular architecture and mechanism of the anaphase-promoting complex** *Nature* 2014, **513**:388–393.

34. Heyman J, De Veylder L: **The anaphase-promoting complex/cyclosome in control of plant development.** *Mol. Plant* 2012, **5**:1182–1194.

35. Sridhar V V, Surendrarao A, Gonzalez D, Conlan RS, Liu Z: **Transcriptional repression of target genes by LEUNIG and SEUSS, two interacting regulatory proteins for Arabidopsis flower development.** *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**:11494–11499.

36. Pfluger J, Zambryski P: **The role of SEUSS in auxin response and floral organ patterning.** *Development* 2004, **131**:4697–4707.

37. Stahle MI, Kuehlich J, Staron L, von Arnim AG, Golz JF: **YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis.** *Plant Cell* 2009, **21**:3105–3118.

38. Castro A, Vigneron S, Bernis C, Labbe J-C, Lorca T: **Xkid Is Degraded in a D-Box, KEN-Box, and A-Box-Independent Pathway** *Mol. Cell. Biol.* 2003, **23**:4126–4138.

39. Klahre U, Friederich E, Kost B, Louvard D, Chua NH: **Villin-like actin-binding proteins are expressed ubiquitously in Arabidopsis.** *Plant Physiol.* 2000, **122**:35–48.

40. Thomas C, Tholl S, Moes D, Dieterle M, Papuga J, Moreau F, Steinmetz A: **Actin bundling in plants.** *Cell Motil. Cytoskeleton* 2009, **66**:940–957.

41. Tominaga M, Yokota E, Vidali L, Sonobe S, Hepler PK, Shimmen T: **The role of plant villin in the organization of the actin cytoskeleton, cytoplasmic streaming and the architecture of the transvacuolar strand in root hair cells of Hydrocharis** *Planta* 2000, **210**:836–843.

42. Gottesman S: **Proteases and their targets in Escherichia coli**. *Annu. Rev. Genet.* 1996, **30**:465–506.

43. Goto-Yamada S, Mano S, Yamada K, Oikawa K, Hosokawa Y, Hara-Nishimura I, Nishimura M: **Dynamics of the light-dependent transition of plant peroxisomes.** *Plant Cell Physiol*. 2015, **56**:1264–1271.

44. Menges M, Hennig L, Gruissem W, Murray JAH: **Cell cycle-regulated gene expression in Arabidopsis.** *J. Biol. Chem.* 2002, **277**:41987–42002.

45. Ok SH, Jeong HJ, Bae JM, Shin J-S, Luan S, Kim K-N: **Novel CIPK1-associated proteins in Arabidopsis contain an evolutionarily**

conserved C-terminal region that mediates nuclear localization. *Plant Physiol.* 2005, **139**:138–150.

46. Hartmann AM, Nayler O, Schwaiger FW, Obermeier A, Stamm S: **The Interaction and Colocalization of Sam68 with the Splicing-associated Factor YT521-B in Nuclear Dots Is Regulated by the Src Family Kinase p59fyn** *Mol. Biol. Cell* 1999, **10**:3909–3926.

47. Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M: **Selective elimination of messenger RNA prevents an incidence of untimely meiosis.** *Nature* 2006, **442**:45–50.

48. Heyman J, Cools T, Vandenbussche F, Heyndrickx KS, Van Leene J, Vercauteren I, Vanderauwera S, Vandepoele K, De Jaeger G, Van Der Straeten D, et al.: **ERF115 controls root quiescent center cell division and stem cell replenishment.** *Science* 2013, **342**:860–863.

Part IV. Unravelling complex traits by complex identification

Chapter 6. A high-throughput platform for screening yield enhancement genes

6.1 The more the merrier: High-throughput screening of yield enhancement genes

Introduction

The improvement of intrinsic yield qualities in crops is one of the main goals in the agro-industry. Conventional breeding provided significant improvements, but is limited within the boundaries of the species. The use of biotechnological tools enables to break that genetic boundary, but also poses huge challenges. To decide which promoter-gene combination could alter a plant's intrinsic yield characteristics, thorough knowledge is required about where and when growth regulators are crucial in contributing to an enhanced yield phenotype. In the past decades, this was tackled by bluntly empirical testing the effect of changing expression levels of genes in model plants and extrapolating the results to the crops. This is illustrated by two private initiatives. The company Mendel biotech for example tested numerous transcription factors by constitutive overexpression in Arabidopsis. CropDesign developed TraitMill, an automated plant evaluation platform that allows high-throughput testing of the effect of transgenes on yield and other agronomical valuable traits in rice. Rice is considered as an appropriate model for cereal crops such as maize and wheat with its rather small genome (389 Mb) and evolutionary relatedness. These efforts led to the elucidation of numerous individual growth regulators, but knowledge from these scattered data is still far from sufficient to efficiently engineer complex biological traits such as growth or seed yield. Finding the links between individual growth regulators will be key to define which molecular networks play in defining yield traits. A significant part of connections between biological entities is defined by protein-protein interactions (PPI). PPI are indeed known to form the basis of many cellular processes and biological functions. Elucidation of the molecular interactions between yield stimulating proteins, collectively called the 'yield interactome', could therefore gain insight in how molecular networks control complex phenotypes

In the previous chapters, we described the development of a functional interactomics tool that enables screening for protein complex data in different rice tissues. For this, we established a workflow that has the potential to increase the efficiency of TAP as compared to the state of the art [1]. The workflow includes a Gateway® technology adapted cloning toolbox that allows versatile combination of a gene of interest with different promoters and TAP tags, reasonably fast biomass generation, an optimised purification protocol based on the GS and GSgreen tag and ultrasensitive mass spectrometry. The building blocks we selected based on experiments using the cell cycle protein CKS1 were the Cauliflower Mozaic Virus 35S (PRO_{35S}) and the rice *GOS2* promoter (PRO_{GOS2}), along with the GS and GSgreen tag.

We now employed this technology for screening 33 genes that showed to have an effect on growth or seed yield in rice upon perturbation. The majority (25) of the

genes was tested in rice callus, but we also further explored screening PPI from shoots immediately regenerated from T0 shoots (10 genes) and T1 seedlings (5 genes). The purpose is twofold; first we wanted to mine the protein interaction networks around these 33 genes and second, the assaying of a large amount of different genes allowed to make an estimate on the performance of the retained TAP tools in rice.

Overview of the platform

Selection and cloning of the baits

We screened a total of 33 additional genes for protein-protein interactions in different rice tissues. Indications from literature or *in house* phenotypic screens of overexpression or down-regulation constructs hinted to involvement of these genes in the regulation of growth or seed yield processes. To challenge the versatility of the TAP platform towards different kinds of protein baits, we chose the genes as such that they vary in biological function or protein type. To further consolidate previous results with the CKS1 test bait in finding optimal TAP tools for use in rice, we fused the bait once with the GS tag and once with the GSgreen tag. We ensured that the two tags were equally tagged as N-terminal and C-terminal fusion, by alternating the fusion orientation of the tags across all baits. In general, GS-tagged constructs were combined with the PRO_{35S} and GSgreen-tagged constructs by the PRO_{GOS2}. We opted to combine the GSgreen tag with a less strong expressing promoter, since significant decay of the tag was visible with the PRO_{35S} driven CKS1 construct and we were afraid that the GFP breakdown product could interfere with complex purification. Therefore we opted for lowering the bait's expression level and hence also lowering the decay product.

Overview of the experimental settings

Purifications were performed based on the earlier described experiments. Briefly, for callus purifications, 15g of callus was used for 2 replicate purifications. A total protein input of 50mg was used per purification from GS-tagged proteins. For GSgreen-tagged proteins, only 25mg input was applied, because the affinity matrix had a significantly lower binding capacity (7.5 µg for GFP trap agarose compared to 50µg for IgG sepharose for 25µL effective beads). A large excess of beads or capture antibody in relation to protein extract should indeed be avoided, to keep the amount of background binding proteins as low as possible [2]. The protein extract was incubated with IgG sepharose or GFP-trap agarose affinity beads for GS- and GSgreen-tagged baits respectively. Bait was subsequently eluted by proteolytic cleavage catalysed by the Tobacco Etch Virus (TEV) protease at 16°C for 1 hour. After a second enrichment step using streptavidin sepharose beads, the complexes were eluted by addition of 40µL of NuPAGE sample buffer containing 20mM desthiobiotin. After a short NuPAGE run, samples were digested *in-gel* by trypsin before mass spectrometry analysis using a nano LC-coupled Velos LTQ Orbitrap system. For purifications on plant tissues (seedlings and shoots), the same purification protocol was applied, with the exception of the first affinity step and the mass spectrometer used. There, the protein extracts (ideally in the range of 150mg total protein per experiment) from the plant cells were applied on affinity beads immobilised on a column, connected with a peristaltic pump. The extract was applied multiple (3-4) times over the column to ensure high contact ration of

extract with the beads. The samples were then analysed on the Q Exactive instead of the Velos LTQ Orbitrap system.

A well-known challenge associated with any AP-MS experiment is the discrimination of *bona fide* interactors from non-specific background associations. We gave preference to a subtraction list containing potential background binders based on the principle that contaminating proteins re-occur over independent experiments. This strategy was already successfully applied in the analysis of 100 human protein complexes [3] and was recently introduced for background filtering from AP-MS experiments in Arabidopsis [4]. We built our list of nonspecific proteins based on the occurrence of proteins in all TAP read-outs from both negative control purifications with wild type samples and from TAP-MS samples, representing 174 experiments in total. Some caution regarding the approach is advised, since this way of filtering does not take into account whether the baits are expected to function in the same pathway or biological process and thus might share some common interactors. We therefore assigned the baits to 34 'bait classes', based on shared membership in the same protein complex, the same gene family [5] or the same biological process according to their gene ontology annotation [6]. We set an arbitrary threshold as such that proteins present in two or more different bait groups were considered nonspecific. This resulted in a list of 683 potential background and non-specific interactors that we removed from our interactor lists. Evidently, this type of background filtering gains in discriminative power with an increasing number of experiments. Due to the yet limited (60 experiments from 13 bait groups representing 16 different proteins) experiments supporting a frequent binders list from plant experiments, we set the cut-off for assigning an interactor as nonspecific in this case to proteins present in more than one different bait group. Here we obtained a list of 951 potential background proteins. Besides the obvious problem that the frequency filters are arbitrarily chosen, it is possible that a true high abundance interactor for a given bait is also detected in lower abundance with other baits in negative control runs. For these reasons, obtaining a quantitative measure for the presence of the given hit or prey protein across all purifications may assist in determining the likelihood that the interaction is indeed significant [4,7].

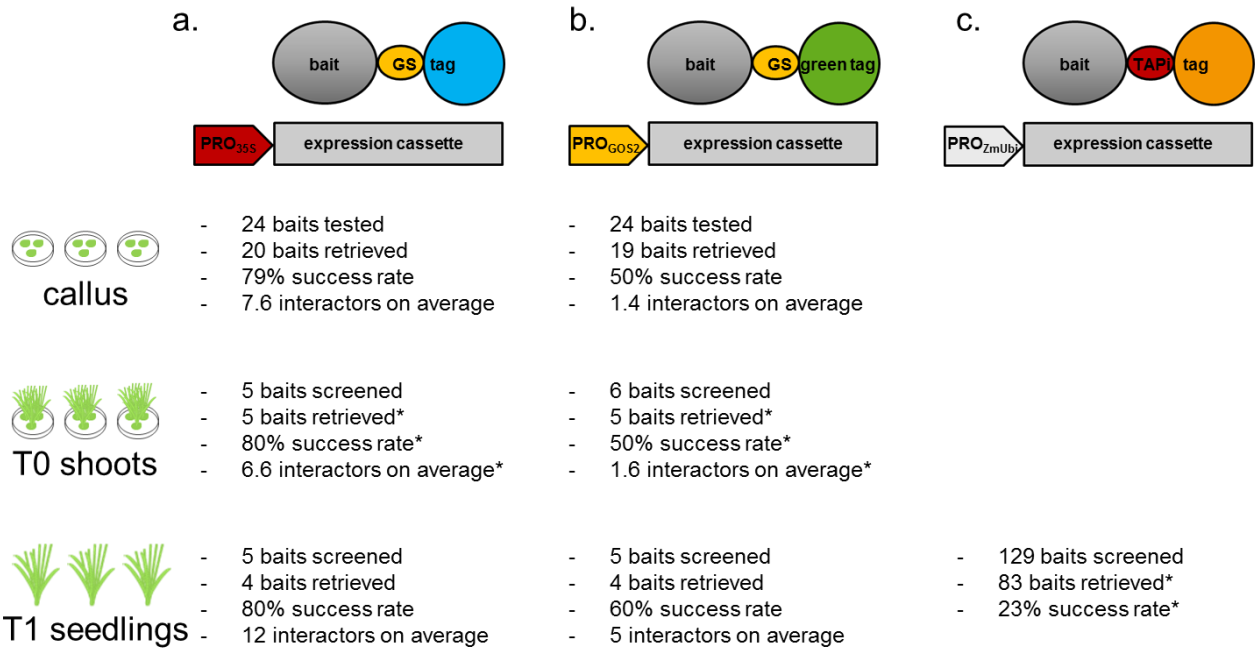
Another crucial parameter in verifying which proteins are 'true' binders is reproducibility. In general, interactions identified in at least two independent purifications are considered as high confidence [8]. Proteins co-purified in only one experiment are in that case mentioned in an additional 'low-confidence' dataset, representing potential interaction partners that are only weakly associated with the complex and requiring additional data to confirm genuine association. In the following sections, proteins will be considered a potential interactor if it was found in at least two independent experiments unless otherwise stated.

Overview of the callus platform

Generation of transgenic callus expressing the bait proteins is the most straightforward way to produce relatively fast an in principle unlimited amount of protein extract. Twenty-five baits were screened in callus tissue. For three of them, we were not able to retrieve the bait in any fusion construct. These bait proteins might be subjected to heavy protein-turnover, or their function might be

compromised by the addition of the bulky tags. For three additional baits, only one fusion resulted in successful retrieval of bait protein. In this case, we hypothesise that one specific fusion resulted in the bait construct being recognised as 'aberrant' and therefore targeted for degradation by the cellular machinery. Summarised, for 25 bait proteins, we constructed 50 TAP constructs, 41 of which were purified in amounts sufficient for their identification by mass spectrometry (Figure S1). As mentioned earlier, most of the bait proteins were tested in a TAP construct containing PRO_{35S} and the GS tag and in a construct with the PR_{GOS2} and the GSgreen tag. For 19 of the 20 GS-tagged constructs that were sufficiently expressed for detection after AP-MS, we were able to isolate complexes with one or more interacting proteins. This boils down to a success rate of 79% and an average retrieval of 7.6 interactors per construct (Figure 1a). Nineteen of the 25 GSgreen-tagged constructs accumulated sufficiently for retrieval in the MS read-out after purification. From these 19, 12 resulted in the co-precipitation of interactors. Hence, the success rate (50%) and the average number of interactors retrieved per construct (1.4) was significantly lower compared to GS purifications (Figure 1b). Two factors could contribute to this significant lower success rate. First, the initial input from GSgreen purifications is only half of the input from GS purifications. This could result in sub-optimal levels of final eluate for mass spec analysis. Second, the PRO_{GOS2} driving the expression of the GSgreen-tagged constructs results in lower protein levels than constructs under control of the PRO_{35S}. Possibly this hampers optimal competition of the bait with its endogenous counterpart in the wild type background for incorporation in the complexes, resulting in accumulation and ultimate purification of mainly empty bait. We challenged the first possibility by repeating GSgreen purification for 4 baits, this time using 50mg as total protein input and observed a trend of recovering more interacting proteins (Figure S2). The effect of the promoter on the bait was tested in additional experiments with CDKD and APC10 (Figure 2). The two proteins were tested as a GS^{rhino}-tagged fusion protein in combination with both PRO_{35S} and PRO_{GOS2}. AP-MS experiments showed no significant differences between the two promoters, leading to the conclusion that the low success rate from GSgreen-tagged constructs was due to a too low biomass input.

Figure 1. Comparison of the performance parameters for different constructs and platforms.



Performance parameters for screening 33 bait proteins are shown for callus, T0 shoots and T1 seedling platforms for GS (a) and GSgreen (b) constructs described in this work and for the platform described by Rohila *et al.* [1] (c). The asterisk for performance data from Rohila *et al.* [1] and T0 shoots stresses the fact that this is related to nonreplicate experiments. The orientation of the TAP tags in the figure do not necessarily hold true for the actual constructs tested.

Figure 2. Comparison of promoter performance using GS^{rhino}-tagged APC10 and CDKD.

			callus		T0 shoots		T1 seedlings	
accession	description	symbol	PRO_{GOS2}	PRO_{35S}	PRO_{GOS2}	PRO_{35S}	PRO_{GOS2}	PRO_{35S}
loc_os05g32600	CYCLIN-DEPENDENT KINASE D 1;3	CDKD1;3	2	2	1	1	2	2
loc_os03g52750	CYCLIN H 1	CYCH1	2	2	1	1	2	2
loc_os11g28350	CAK assembly factor MAT1	MAT1	2	2	1	1	2	2
loc_os06g07480	CAK assembly factor MAT1	MAT1	2	2	1	1		2
loc_os05g05260	XERODERMA PIGMENTOSUM D	XPD	2	2	1	1	2	2
loc_os02g03340	TFIIH p34 subunit		2	2	1	1	2	2
loc_os04g42990	TFIIH p44 subunit		2	2	1	1	2	2
loc_os04g58350	TFIIH p52 subunit		2	2	1	1	2	2
loc_os08g25060	TFIIH p62 subunit		2	2	1	1	2	2
loc_os06g22820	CDKF-1/CAK1AT	CDKF1		1				

			callus		T0 shoots		T1 seedlings	
accession	description	symbol	PRO_{GOS2}	PRO_{35S}	PRO_{GOS2}	PRO_{35S}	PRO_{GOS2}	PRO_{35S}
loc_os05g50360	ANAPHASE PROMOTING COMPLEX SUBUNIT 10	APC10	2	2	1	1	2	2
loc_os06g41750	ANAPHASE PROMOTING COMPLEX SUBUNIT 3	APC3	2	2	1	1	2	2
loc_os03g13370	ANAPHASE PROMOTING COMPLEX SUBUNIT 6	APC6	2	2			1	1
loc_os05g05720	ANAPHASE PROMOTING COMPLEX SUBUNIT 7	APC7	2	2	1	1	1	2
loc_os02g43920	ANAPHASE PROMOTING COMPLEX SUBUNIT 8	APC8	2	2	1	1	2	2
os05g0354300	ANAPHASE PROMOTING COMPLEX SUBUNIT 1	APC1	2	2	1	1	2	2
loc_os02g54490	ANAPHASE PROMOTING COMPLEX SUBUNIT 4	APC4	1	2		1	1	
loc_os12g43120	ANAPHASE PROMOTING COMPLEX SUBUNIT 5	APC5	1	2	1	1	2	2
loc_os04g40830	ANAPHASE PROMOTING COMPLEX SUBUNIT 2	APC2	2	2	1	1	2	2
loc_os03g19059	ANAPHASE PROMOTING COMPLEX SUBUNIT 11	APC11		2				
loc_os03g03150	CELL CYCLE SWITCH PROTEIN 52A 1	CCS52A1		1				2
os10g0575950	SAMBA	SAMBA		2				

From interaction datasets using PRO_{35S}- and PR_{GOS2}-driven APC10 and CDKD baits, only the core complex constituents are shown. Numbers illustrate how many times the protein was retrieved. Purifications using callus and T1 seedling biomass were performed twice per construct, purifications using T0 shoots only once.

Exploration of the purification strategy in plant tissues

Some proteins are discretely expressed or associate with proteins that are discretely expressed. In that case, screening interactions from callus tissue will result in false negatives, since the appropriate development or anatomical context is not in place. In that case, the complex can be purified from plant samples. We opted to sample the plants in an early stage to have a high ratio of proliferative cells with the underlying assumption that genes controlling growth processes are active in this phase. By regenerating shoots from callus, we could obtain this type of tissue in reasonable time. We tested 14 constructs comprising 10 baits and found 13 of them sufficiently expressed for detection after TAP-MS (Figure S3). Experiments were only performed once, because the high dependency on regeneration efficiency compromised obtaining sufficient biomass to provide protein input for two replicate purifications. This is of course problematic for assessing the performance and reproducibility of the method. Alternatively, T1 seeds were harvested from transgenic lines expressing the TAP constructs. From these seeds, we cultivated two week old seedlings. These were grown on selection medium to favour growth of homozygous over heterozygous plants and eliminate null-segregants. From 550 plants, we harvested on average 20g of biomass, providing 300mg of total protein. This was sufficient for two replicate experiments using 150mg total protein each. Here, we screened 5 baits using 10 constructs that were also tested in callus (Figure S4). For 1 of the 5 baits we failed in accumulating sufficient protein levels for retrieval in the MS read-out after purification for both the GS- and GSgreen-tagged fusion construct. The other 4 baits did result in successful bait recovery. In the case of GS-tagged constructs all four baits co-purified proteins (80% success rate), with an average of 12 retrieved interactors (Figure 1a). For GSgreen-tagged constructs, we found interactors in three of the four cases (60% success rates) and obtained an average of 5 interactors (Figure 1b). Since the five proteins screened in seedlings were also tested in callus, we were able to compare the proteins that were identified (data not shown). The significant amount of proteins uniquely observed in seedlings for some of the baits might hint to proteins that are only present in a developing context. However, it is important to take into account that, with in total 15 proteins screened from plant tissues, the background list is at this stage not sufficiently reflecting which proteins might re-occur over independent experiments. Therefore a large part of the 'plant-specific' interactors could also represent false positives.

Conclusion from the high-throughput platform

We further elaborated the integrated workflow developed from pilot experiments using CKS1, CDKD and APC10 to assay the throughput of baits and screened 25 proteins in rice callus, 9 proteins in shoots regenerated from callus, and 5 proteins from seedlings grown from seeds from selected transgenic lines. In rice, the major model for cereal crops, there is only one exhaustive TAP-MS study to date to our

knowledge. This study, reported over 2 publications, encompasses the unravelling of kinase complexes from rice plants in nonreplicate experiments. They found at least one interaction partner for 23% of the 129 kinases screened [1,9]. Kinases are a tricky class of proteins to screen interactors from though. This, and the fact that they only tested N-terminal fusions, might have compromised the true potential of their platform. Our workflow, when compared with this previous method, has numerous favourable properties. First, it significantly saves resources, labour and time for the systematic generation of cell or seed lines for expression of affinity-tagged bait proteins. Next, it has higher sensitivity and thus minimises sample amounts required for successful isolation and analysis of protein complexes. And last, it generates protein interaction data with increased success rates, average amounts of interactors and reproducibility. For fair comparison, we retrieved interactors for all 4 of the baits annotated as protein kinase (Figure S5). As these were mainly screened in rice callus, the increased success rate for GS purifications could also (partially) be attributed to the use of this more straightforward biological system.

In the presented method, we used MultiSite Gateway®-mediated recombination of promoter, gene of interest and TAP tag for rapid generation of TAP expression vectors. This takes advantage of the versatility of the Gateway® system to combine any gene with the desired promoter and TAP tag in both N- or C-terminal fusions. Starting from a cDNA entry clone, sufficient callus cells expressing the bait are reliably obtained within 3.5 months. A similar timeframe is required to generate shoot material, which is significantly less than the time needed for generating plant material with previous methods. The dependence on regeneration efficiency is however a major trade-off, preventing to provide amounts sufficient for duplicate experiments. This hampers the assessment of reproducibility. We therefore also tested the use of two weeks old T1 seedlings, prolonging the timeline to 10.5 months.

Another bottleneck in the previous TAP analysis is linked to the limited overall sensitivity of the workflow, requiring starting material in the range of 50-60g material. This in turn poses huge pressure on available resources and experimental logistics for high throughput study. Overall sensitivity in TAP experiments is determined by both the purification procedure and the mass spectrometric analysis. We introduce a purification strategy based on the GS tag and the related GSgreen tag, in combination with the ultrasensitive Orbitrap MS. The GS tag was shown to allow efficient double-affinity with higher specificity and higher yields compared to the classical tag [10,11]. The GSgreen tag is basically similar to the GS tag with a GFP portion replacing the tandem repeat of the ZZ domain of protein G. The combined purification approach with Orbitrap mass spectrometry allows the identification of specific complex partners from a minimal protein input of 50mg per purification. We tested halving the protein input with GSgreen purifications, but observed a significant decrease in success rate.

To generate a reliable dataset of potential interactors, a clear estimation of the false-positive interactions is required. First, the GS and GSgreen tag allow efficient removal of the bulk of contaminants during the two-step purification process. Next, we built a contaminant dataset based on occurrence over independent purifications

of unrelated bait groups to efficiently remove the remainder of the contaminants. Importantly, there are drawbacks to this type of heuristic management of the interaction data. First, frequency filters are only really applicable to larger scale studies, but not to studies of individual proteins or small numbers of baits [12]. This is illustrated in our dataset of plant purifications. The limited set of experiments forced us to also consider callus experiments and required the filter to be set very stringently. Second, the frequency filters are often chosen in a context or data-dependent manner, and are hard to apply if baits are functionally related and copurify with similar sets of proteins. To address this issue, we grouped our baits according to co-membership in protein complexes, gene families of biological processes and considered the resulting 34 bait groups for applying the frequency filters. In addition, the filtering method is only based on binary data (presence or absence of the interactor in each bait dataset), and ignore quantitative features of the mass spectrometry results. Genuine abundant interactors of any given bait are often missed since they are occasionally identified as low abundance interactors with several other baits. Quantitiave measures of protein abundance in interaction data sets are therefore important to resolve whether the interaction bait-prey is specific [4,7]. Another quality parameter in TAP-MS experiments is reproducibility. To offer the best quality interaction dataset, we only considered interactions confirmed in at least two replicate experiments as *bona fide*.

The resulting interaction datasets of the 33 baits screened revealed their local protein interactions providing a peak insight into the gene's potential functioning. The acquired knowledge enables to make more targeted choices when selecting genes for testing as candidate yield-improvement gene. Also, in regard to the guilt-by-association concept, interaction partners from yield-enhancing genes might have the same yield stimulating characteristic. And third, the initial bait and the interaction partner can be combined in a co-expression construct, allowing additive or synergistic effects. These factors will help to increase the success rate of successfully identifying optimal transgenes for enhancing yield phenotypes.

Our callus platform supports screening unknown proteins as it enables easy and fast assaying for interactors. Undifferentiated cells are indeed an excellent hypothesis-generating tool and cover a wealth of basic cellular processes, such as cell cycle regulation, hormone signalling and intracellular trafficking. For proteins involved in processes related to specific stages in development, such as for example light responses, circadian clock or photosynthesis, whole plants rather than cultured cells should be used for screening interactors. Low abundant proteins or proteins that show a very discrete expression pattern throughout development, might require specific sampling of the tissues where and when the protein of interest is expressed. In that light we provided a proof of concept by assaying the interaction partners of APC10 and CDKD from the proliferation zone of the leaf (cfr. Chapter 5). A recent, more elaborated study in corn showed that TAP purification from dissected tissues enables studying dynamics of low-abundance protein complexes throughout the growth zone in a developing organ [12].

Whereas the lower performance of the GSgreen tag in callus purifications could (partially) be attributed to a lower protein input, this does not hold true for

purifications from plant tissues. Therefore we suggest to use the GS tag for future purifications. As this underperforming of the GSgreen tag might be caused by a lower binding capacity we further recommend re-evaluating GSgreen purifications if more performant affinity matrices for GFP enrichment would become available. Comparison of the PRO_{GOS2} and PRO_{35S} using three different bait proteins (CKS1 described in chapter 4, APC10 and CDKD described in this chapter) resulted in similar purification yields. It is therefore advisable to retain the same promoter as for which a yield effect was observed when screening for interactors.

In conclusion, we developed an improved functional interactomics approach that sustains high-throughput screening for protein interactions from a portfolio of rice tissues. Interaction datasets generated from this workflow will be key in systems-oriented efforts to understand the biological processes behind plant growth or seed yield. It is indeed the only method to retrieve global information on how proteins are organized into systems of functionally interacting complexes. We demonstrated the potential of our workflow by screening 33 growth and/or seed yield-related genes in different rice tissues and will discuss the interaction datasets from 3 of these gene in more detail in the upcoming chapters.

Materials & Methods

Generation of TAP-compatible ENTRY vectors

Open Reading Frames from the genes under investigation were amplified from expression vectors that contained the sequence and previously were used in the TraitMill platform. Primers were designed using the VectorNTI primer design tool with the att-sequences added.

For generating amplicons for N-terminal tagging, primers were selected as such that no Kozak sequence was picked up. PCR was performed using the Phusion Hot Start II high-fidelity DNA polymerase kit (Thermo Fisher Scientific, Waltham, MA). To verify amplification, PCR product was loaded on a 1% agarose gel and run for 25 min at 100V. Subsequently, fragments of the corresponding size were excised from gel and extracted using the QIAquick gel extraction kit (Qiagen, Venlo, The Netherlands). Next, the amplicon was cloned in the compatible pDONR™P2R-P3 vector (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Successful recombination was verified by restriction digest analysis and sequencing of the Gateway® cassette.

For C-terminal tagging, primers were selected as such that no stop codon was present in the amplicon. PCR was performed using the Phusion Hot Start II high-fidelity DNA polymerase kit (Thermo Fisher Scientific, Waltham, MA). To verify amplification, PCR product was loaded on a 1% agarose gel and run for 25 min at 100V. Subsequently, fragments of the corresponding size were excised from gel and extracted using the QIAquick gel extraction kit (Qiagen, Venlo, The Netherlands). Next, the amplicon was cloned in the compatible pDONR™221 vector (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Successful recombination was verified by restriction digest analysis and sequencing of the Gateway® cassette.

Gateway® cloning to generate TAP expression vectors

Recombination of the promoter, gene sequence and TAP tags was performed using the standard MultiSite Gateway® cloning technology to generate TAP destination vectors.

The resulting TAP expression vectors were verified by restriction digest analysis and transferred to *Agrobacterium tumefaciens* lba4044/pal4404 by electroporation. Transformed bacteria were selected on yeast extract borth medium with the appropriate antibiotics and verified by colony PCR using Takara Taq polymerase (Takara Bio Inc, Shiga, Japan) according to the manufacturer's instructions.

Transformation and biomass generation

Agrobacterium tumefaciens mediated transformation of *Oryza sativa* (ecotype japonica) seeds was performed according to patent WO2001006833 A1 with minor modifications. After mechanical dehusking using rice husker Kett US TR120, 150-200 seeds were surface sterilized with 6% sodium hypochlorite solution for 45 minutes and washed with sterile water. Afterwards, seeds were transferred to induction medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 4 g/L gelrite, 2 mg/L 2,4-D) and allowed to germinate at 32°C under continuous light of 3000 lux. Six days after germination, the seeds were briefly submerged in liquid infection medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 68.5 g/L sucrose, 36 g/L D+ glucose-monohydrate, filter sterilised) containing 100 µM acetosyringone and transgenic A. tumefaciens lba4404/pal4404 containing the TAP destination vector (OD600 0,05-0,1) and transferred to co-cultivation medium (pH 5.2, 4 g/L MS salts, 1 mL/L MS vitamins, 300 mg/L CasaminoAcids, 30 g/L sucrose, 10 g/L D+ glucose-monohydrate, 4 g/L gelrite, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 µM acetosyringone). Co-cultivation was allowed for three days at 25°C in darkness. Thereafter, the explants were removed from the seeds, washed with 250 mg/L cefotaxime and transferred to selection medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 2878 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 7 g/L agarose type 1, 2 mg/L 2,4-dichlorophenoxyacetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 35 mg/L G418 disulfate) for incubation under continuous light (3000 lux) at 32°C. Twelve days later, microcalli were isolated and transferred onto fresh selection medium, refreshed every ten days, and grown until 30 g of callus was obtained. The callus material was then harvested in liquid nitrogen and stored at -80°C for subsequent analysis.

Transformation and T0 shoot generation

For the generation of T0 shoot material, the same transformation protocol was utilized as described above. After isolation, the microcalli were transferred onto pre-regeneration medium (pH 5.8, 4 g/L MS salts, 1 mL/L MS vitamins, 500 mg/L L-Proline, 300 mg/L CasaminoAcids, 30 g/L sucrose, 7 g/L agarose type 1, 2 mg/L kinetin, 1 mg/L α-naphthalene acetic acid, 5 mg/L abscissic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 20 mg/L G418 disulfate) and incubated for one week at 32°C under continuous light (3000 lux). Resistant callus was first brought to regeneration medium I (pH 5.8, 4.3 g/L MS salts, 1 mL/L MS vitamins, 2 g/L CasaminoAcids, 30 g/L sucrose, 30 g/L sorbitol, 10 g/L agarose type 1, 2 mg/L kinetin, 0.02 mg/L α-naphthalene acetic acid, 100 mg/L cefotaxime, 100 mg/L

vancomycin, 20 mg/L G418 disulfate) for seven days and then transferred to regeneration medium II (pH 5.8, 4.3 g/L MS salts, 1 mL/L MS vitamins, 2 g/L CasaminoAcids, 30 g/L sucrose, 30 g/L sorbitol, 7 g/L agarose type 1, 2 mg/L kinetin, 0.02 mg/L α-naphthalene acetic acid, 100 mg/L cefotaxime, 100 mg/L vancomycin, 20 mg/L G418 disulfate) for two additional weeks at 32°C under continuous light (lux 3000). Plants whose shoot and root grew more than 1 cm in length were ultimately transferred to growth medium (pH 5.8, 2.15 g/L MS medium micro and macro, 0.5 mL (0.5 x) B5 vitamins, 10 g/L sucrose, 0.05 mg/L α-naphthalene acetic acid, 0.75 g/L $MgCl_2.6H_2O$, 2.5 g/L gelrite) and incubated for two weeks at 32°C under continuous light (3000 lux) before being harvested in liquid nitrogen and stored at -80°C for subsequent analysis.

Transformation and T1 seedling generation

Transgenic seeds containing the TAP expression vector were derived from the transformation protocol generating shoots described above. Instead of harvesting, 60 plants, each derived from an individual transformation event, were transferred to the greenhouse and further grown until seeds could be harvested. For growing seedlings, seeds were first dehusked and sterilized as previously described before sowing them on growth medium containing the selective agent (pH 5.8, 2.15 g/L MS medium micro and macro, 0.5 mL (0.5 x) B5 vitamins, 10 g/L sucrose, 0.05 mg/L α-naphthalene acetic acid, 0.75 g/L $MgCl_2.6H_2O$, 7 g/L agarose, 20 mg/L G418 disulfate). T1 seedlings were grown in a growth chamber under short day conditions at 32°C, and harvested two weeks after sowing in liquid nitrogen and stored at -80°C for subsequent analysis.

Expression analysis of the bait proteins

Material was ground to homogeneity in liquid nitrogen with mortar and pestle. About 200 µL of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM MgCl2, 150 mM NaCl, 15 mM pNitrophenyl phospate, 60 mM β- glycerophosphate, 0.1% NP-40, 0.1 mM Na3VO4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTAfree Ultra Complete tablet (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) was added and homogenized with a 1.5-mL pellet mixer. Homogenized samples were flash frozen in liquid nitrogen, thawed on ice and centrifuged twice for 15 min at 4°C at 20,800 g. Protein concentrations were determined by Bradford assay (Bio-rad, Hercules, CA). Fifty µg of total protein extract was loaded for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 0.75 mm 12% Mini-PROTEAN® TGX™ precast gels (Bio-Rad, Hercules, CA) for 20 min at 300 V in TGX running buffer (25 mM Tris-HCl, pH 8.3, 1.92M glycine, 35 mM SDS). Resolved proteins were transferred to PVDF membranes using Trans-Blot® Turbo™ Mini PVDF transfer packs and the Trans-Blot® Turbo™ Transfer system (Bio-rad, Hercules, CA) according to instructions of the manufacturer. Blotted PVDF membranes were then incubated in blocking buffer (3% Difco™ skimmed milk (w/v) in TBS-T buffer (50 mM Tris, 150 mM NaCl pH8.0, 0.1% Triton X-100)) overnight at 4°C or 1 h at room temperature (RT) on an orbital shaker. After this blocking step, membranes were incubated for 1 h at RT with peroxidase anti-peroxidase antibody (Sigma-Aldrich, Saint-Louis, MO) in blocking buffer on an orbital shaker. Membranes were washed 1 x 15 min and 4 x 5 min with TBS-T buffer. Bound antibody was detected by mixing equal amounts of the two chemiluminescent reagents from the ECL-kit (Perkinelmer, Waltham, MA) and

incubating for 1 min. Membranes were placed in a film cassette and exposed to an Amersham hyperfilm™ ECL film (GE Healthcare, Wauwatosa, WI) in a dark room, where autoradiograms were also developed.

Protein extraction for TAP experiments
Callus material was ground to homogeneity in liquid nitrogen with mortar and pestle. Plant material (T0 shoots, T1 seedlings) was ground to homogeneity in liquid nitrogen with a hand blender (Braun GmbH, Kronberg, Germany). Crude protein extracts were prepared in two volumes of extraction buffer (25 mM Tris-HCl pH 7.6, 15 mM $MgCl_2$, 150 mM NaCl, 15 mM *p*-nitrophenyl phosphate, 60 mM β-glycerophosphate, 0,1% NP-40, 0.1 mM Na_3VO_4, 1 mM NaF, 1 mM PMSF, 1 µM E64, EDTA-free Ultra Complete tablet Easypack (1/10 mL) (Roche Diagnostics, Brussels, Belgium), 5% Ethylene glycol) at 4°C using an Ultra-Turrax T25 mixer (IKA Works, Wilmington, NC). Soluble fraction was obtained from isolating the supernatans after double centrifugation at 36,900 g for 20 min at 4°C. The extract was passed through four layers of miracloth (Merck KGaA, Darmstadt, Germany) and kept on ice.

GS TAP purifications
For purifications from callus cells, purifications were performed as described by Van Leene et al. (2010) [8] with minor modifications. The protein extract was added to 25 µL of effective immunoglobulin G (IgG) Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 375 µL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 50 µL wash buffer and 2x 10U (2 x 1 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 100 µL wash buffer was added to the beads to collect residual eluate. The resulting eluate was incubated for 1 h at 4°C under gentle rotation with 25 µL effective streptavidin Sepharose High Performance beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL wash buffer. Streptavidin beads were transferred to a mobicol column and washed with 100 column volumes or 2.5 mL wash buffer. Complexes were eluted in 40 µL NuPAGE sample buffer containing 20 mM desthiobiotin (Sigma-Aldrich, Saint-Louis, MO) by 5 min incubation on ice, followed by centrifugation at 1,500 rpm at 4°C.

For seedling and shoot material, the first affinity purification step was performed on a Poly-Prep column (Bio-Rad, Hercules, CA) containing 100 µL of effective IgG Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 1 mL extraction buffer, with a peristaltic pump (GE Healthcare, Wauwatosa, WI), at flow rate 1 mL/min. The Poly-Prep column was then mounted to a two-way valve in a vacuum manifold system for washing with 150 mL or 15

column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 200 µL wash buffer and 2x 40U (2 x 4 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 400 µL wash buffer was added to the beads to collect residual eluate. The rest of the protocol proceeded similarly as described above.

GSgreen TAP purifications

The protein extract was added to 100 µL of effective GFP-Trap® agarose beads (Chromotek GmbH, Planegg-Martinsried, Germany), pre-equilibrated with 3x 1 mL extraction buffer. After incubation for 1 hour at 4°C under gentle rotation, the beads were transferred to a Poly-Prep column (Bio-Rad, Hercules, CA) mounted to a two-way valve in a vacuum manifold system (Grace, Columbia, MD) and washed with 15 mL or 150 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). Bound complexes were eluted by digestion in a mobicol column (MoBiTec GmbH, Göttingen, Germany) using 200 µL wash buffer and 2x 40U (2 x 4 µL, second boost after 30 min) AcTEV protease (Life Technologies, Carlsbad, MA) for 1 h at 4°C on a shaker. Eluate was collected by two consecutive spinning steps of the mobicol column (MoBiTec GmbH, Göttingen, Germany) in a 2 mL Eppendorf tube for 30 sec at 1,500 rpm at 4°C. In between, 400 µL wash buffer was added to the beads to collect residual eluate. The resulting eluate was incubated for 1 h at 4°C under gentle rotation with 25 µL effective streptavidin Sepharose High Performance beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 250 µL wash buffer. Streptavidin beads were transferred to a mobicol column and washed with 100 column volumes or 2.5 mL wash buffer. Complexes were eluted in 40 µL NuPAGE sample buffer containing 20 mM desthiobiotin (Sigma-Aldrich, Saint-Louis, MO) by 5 min incubation on ice, followed by centrifugation at 1,500 rpm at 4°C.

For seedling and shoot material, the first affinity purification step was performed on a Poly-Prep column (Bio-Rad, Hercules, CA) containing 100 µL of effective IgG Sepharose 6 Fast Flow beads (GE Healthcare, Wauwatosa, WI), pre-equilibrated with 3x 1 mL extraction buffer, with a peristaltic pump (GE Healthcare, Wauwatosa, WI), at flow rate 1 mL/min. The Poly-Prep column was then mounted to a two-way valve in a vacuum manifold system for washing with 150 mL or 15 column volumes wash buffer (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 µM E64, 1 mM PMSF, 5% Ethylene glycol). The rest of the protocol proceeded similarly as described above.

Sample preparation

Purified protein samples were loaded and separated with a short 7-min run on a precast 4-12% gradient NuPAGE Bis-Tris gel (Life Technologies, Carlsbad, CA), fixed in 50% EtOH/2% H3PO4 and visualized with colloidal Coomassie Brilliant Blue G-250 (Sigma-Aldrich, Saint-Louis, MO) staining.

NuPAGE gel containing purified protein samples was destained twice in HPLC-grade water (Thermo Fisher Scientific, Waltham, MA) for 1 h and incubated in 25 mL of reducing buffer (6.66 mM DTT plus 50 mM NH4HCO3 in HPLC-grade water) for 40 min to reduce the polypeptide disulphide bridges. Subsequently, thiol groups were alkylated by incubating the gel for 30 min in 25 mL of alkylating buffer (55 mM iodoacetamide, 50 mM NH4HCO3 in HPLC-grade water) in the dark before washing with HPLC-grade water. The zone containing the protein sample was sliced from the gel and sectioned into different gel plugs. These were washed twice with 600 µL of HPLC-grade water and dehydrated in 600 µL 95% acetonitrile twice for 10 min. The dehydrated gel plugs were submerged and rehydrated in 90 µL trypsin digest buffer (12.5 µg/mL trypsin (MS gold; Promega, Madison, WI) in 50 mM NH4HCO3 and 10% (v/v) acetonitrile in HPLC-grade water) for 30 min at 4°C. Afterwards, trypsin digestion was allowed for 3.5 h at 37°C. Resulting peptide samples were sonicated for 5 min in a sonication bath and the solution covering the gel plugs (containing trypsinized peptides) were kept aside. Remaining gel plugs were completely dehydrated in 95% acetonitrile for 10 min and the remaining acetonitrile solution was added to the first fraction of trypsin digests. The resulting trypsin-digested sample was completely dried in a SpeedVac for 2-3 h at 4°C.

Mass spectrometry analysis using the LTQ velos Orbitrap
Samples from TAP purifications on callus tissues were analysed using a LTQ Velos Orbitrap mass spectrometer.

A nano LC system (NanoLC Ultra 2D system, Eksigent, Dublin, CA) was connected to an LTQ Velos Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) with a trapping column (PepMap 100, C18 precolumn with 5-µm particles, 20mm × 200 µm internal diameter; Dionex), flow rate of 6 µL/min (100% Solvent A) and 5 minutes after injection switched in line with an analytical C18 column (Acclaim PepMap 100, 3-µm particles, 150 mm × 75 µm internal diameter; Dionex). A chip-based nano-electrospray source (TriVersa, Advion Biosystems, Ithaca, NY) operated at 1.8 kV.

Peptides were solubilised in loading Solvent A (2% acetonitrile, 0.1% acetic acid (v/v) in HPLC grade water) and 10 µL of the sample was loaded on the trapping column. Peptide samples were separated with a 65 min gradient at a flow rate of 300 nL/min. MS spectra were recorded in the Orbitrap FT analyzer with a resolution of 60,000 (at m/z 400) and an automatic gain control (AGC) target setting of 500,000. The maximum injection time was set to 500 ms, and lock mass was enabled (polysiloxane ion at m/z 445.12024). Collision-induced dissociation MS/MS spectra were acquired by the ion trap in data-dependent mode, selecting up to the 20 most abundant multiply charged precursor ions from the MS spectrum. The maximum injection time was set to 50 ms and an AGC setting of 7,500. Fragmentation was accomplished by collision- induced dissociation wideband activation at normalized collision energy of 35 eV and an activation time of 30ms. After MS/MS, the m/z precursors were excluded for 30 s.

Mass spectrometry analysis using the Q Exactive
Samples from TAP purifications on plant tissues were analysed using a Q Exactive mass spectrometer.

A nano LC system (Ultimate 3000 RSLC nano, Dionex, Amsterdam, The Netherlands) was set up connected in-line to the Q Exative (Thermo Fisher Scientific, Waltham, MA), with a trapping column (PepMap 100) of 100-µm internal diameter (I.D.) x 20 mm (length) with 5-µm C18 Reprosil-HD beads (Dionex, Amsterdam, The Netherlands), a flow rate 6 µL/min and, 5 min after injection, switched in-line with a reverse phase analytical separating column (Acclaim, PepMap 100) of 75 µm I.D. x 150 mm with 3 µm C18 Reprosil-HD beads (Dionex, Amsterdam, The Netherlands).

Peptides were solubilized in 15 µL of loading Solvent A (0.1% (v/v) trifluoroacetic acid, 2% (v/v) acetonitrile in HPLC grade water), and 5 µL of the peptide sample was loaded on the trapping column. Peptide samples were separated with a 30-min gradient from 2% mobile phase solvent A' (0.1% (v/v) formic acid in HPLC-grade water) to 50% mobile phase solvent B' (0.1% (v/v) formic acid and 80% acetonitrile in HPLC-grade water) at a flow rate of 300 nL/min. The Q Exactive was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition for the ten most abundant peaks in a given MS spectrum. Full-scan MS spectra were acquired in the Orbitrap at a target value of 1E6 with a resolution of 60.000. The ten most intense ions were isolated for fragmentation in the quadrupole, with a dynamic exclusion of 20 seconds. Target value for filling the quadrupole was set to 1E4 ion counts.

Data analysis of mass spectra

Mascot generic files were created with the Mascot Distiller software. Grouping of spectra with a maximum intermediate retention time of 30 s and 0.005-Da precursor tolerance was allowed together with a maximum intermediate scan count of five if possible. No de-isotoping was used and the relative signal-to-noise limit was set to 2. A peak list was generated only when the MS/MS spectrum contained more than ten peaks. The Mascot Daemon interface was used to search peak lists with the Mascot search engine against the rice annotation project (RAP) database [13] and against the Michigan State University rice database (MSU) [14].

Search parameters for data from the LTQ Velos Orbitrap mass spectrometer in Mascot Daemon were as follows: as variable modifications: methionine oxidation and methylation of aspartic acid and glutamic acid, as fixed modifications: carbamidomethylation of cysteines, as mass tolerance on MS: 10 ppm, as MS/MS tolerance: 0.5 Da, as instrument: ESI-TRAP, as peptide charge: of 1+,2+,3+ and as protease: trypsin/P, allowing for one missed cleavage.

Search parameters for data from the Q Exactive mass spectrometer in Mascot Daemon were as follows: variable modifications set to methionine oxidation and methylation of aspartic acid and glutamic acid, fixed modifications to carbamidomethylation of cysteines, mass tolerance on MS 10 ppm, MS/MS tolerance 20 mmu, ESI-QUAD as instrument, 2+ and 3+ as peptide charge and protease trypsin/P, allowing for 1 missed cleavage.

In the Mascot result URL, a Mascot select summary was created with following settings: significance threshold p>0.01, maximum number of hits AUTO, Mudpit scoring, Ion score or expectancy cut-off >0.01, require bold red. Identifications were retained when at least two peptides matched a high confidence rank, with at

least one peptide unique to the protein. The resulting protein identification list was cross-checked against the list of non-specific proteins. Only specific binders were retained in the final list.

Supplementary information

Figure S1. Overview of high-throughput screening in callus

	N-terminal fusion construct			C-terminal fusion construct			
bait n°	construct	bait	interactors	construct	bait	interactors	total interactors
bait 1	PRO_{35S}:GS-bait	2	3	PRO_{GOS2}:bait-GSgr	2	0	3
bait 2	PRO_{35S}:GS-bait	2	0	PRO_{35S}:bait-GS	2	3	3
bait 3	PRO_{35S}:GS-bait	2	4	PRO_{GOS2}:bait-GSgr	2	2	4
bait 4	PRO_{35S}:GS-bait	2	5	PRO_{GOS2}:bait-GSgr	2	4	5
bait 5	PRO_{35S}:GS-bait	1	0	PRO_{GOS2}:bait-GSgr	2	1	1
bait 6	PRO_{GOS2}:GSgr-bait	2	7	PRO_{35S}:bait-GS	2	8	9
bait 7	PRO_{35S}:GS-bait	2	11	PRO_{GOS2}:bait-GSgr	2	6	12
bait 8	PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	2	2
bait 9	PRO_{35S}:GS-bait	2	2	PRO_{GOS2}:bait-GSgr	2	0	2
bait 10	PRO_{35S}:GS-bait	2	1	PRO_{GOS2}:bait-GSgr	2	1	1
bait 11	PRO_{35S}:GS-bait	2	5	PRO_{GOS2}:bait-GSgr	0	0	5
bait 12	PRO_{GOS2}:GSgr-bait	1	0	PRO_{35S}:bait-GS	2	4	4
bait 13	PRO_{GOS2}:GSgr-bait	2	2	PRO_{35S}:bait-GS	2	12	12
bait 14	PRO_{GOS2}:GSgr-bait	2	4	PRO_{35S}:bait-GS	2	5	7
bait 15	PRO_{35S}:GS-bait	2	0	PRO_{GOS2}:bait-GSgr	2	0	0
bait 16	PRO_{GOS2}:GSgr-bait	0	0	PRO_{35S}:bait-GS	0	0	0
bait 17	PRO_{35S}:GS-bait	2	18	PRO_{GOS2}:bait-GSgr	2	2	18
bait 18	PRO_{35S}:GS-bait	1	0	PRO_{GOS2}:bait-GSgr	0	0	0
bait 19	PRO_{35S}:GS-bait	2	5	PRO_{GOS2}:bait-GSgr	2	0	5
bait 20	PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	9	9
bait 21	PRO_{GOS2}:GSgr-bait	2	0	PRO_{GOS2}:bait-GSgr	2	1	1
bait 22	PRO_{GOS2}:GSgr-bait	0	0	PRO_{35S}:bait-GS	0	0	0
bait 23	PRO_{GOS2}:GSgr-bait	2	3	PRO_{35S}:bait-GS	2	78	78
bait 24	PRO_{GOS2}:GSgr-bait	2	1	PRO_{35S}:bait-GS	2	5	5
bait 25	PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	3	3

Overview of the results from screening 25 baits in callus cells. Each bait protein was tagged in an N- and C-terminal fusion construct. The construct for these fusions is specified for each bait in the 'construct' column. In the 'bait' column, the number of times the bait was retrieved is given. We only considered baits as retrieved for a certain construct if these were found in both purifications. The number of interactors shown reflect the interactors that were reproducibly found, i.e. in at least two independent purifications.

Figure S2. Effect of protein input increase for GSgreen purifications

	GSgreen				GS		
			interactors				interactors
bait n°	construct	bait	25mg	50mg	construct	bait	50mg
bait 3	PRO_{GOS2}:bait-GSgr	2	2	2	PRO_{35S}:GS-bait	2	4
bait 4	PRO_{GOS2}:bait-GSgr	2	4	6	PRO_{35S}:GS-bait	2	5
bait 7	PRO_{GOS2}:bait-GSgr	2	6	7	PRO_{35S}:GS-bait	2	11
bait 14	PRO_{GOS2}:GSgr-bait	2	4	4	PRO_{35S}:bait-GS	2	5

Illustration of the effect of increasing the protein input for GSgreen purifications from 25mg to 50mg. As reference, identified interactors for GS purifications are given in the right panel.

Figure S3. Overview of high-throughput screening in T0 shoots

	N-terminal fusion construct			C-terminal fusion construct			
bait n°	**construct**	**bait**	**interactors***	**construct**	**bait**	**interactors***	**total interactors***
bait 4	PRO_{35S}:GS-bait	1	2	not tested	-	-	2
bait 7	PRO_{35S}:GS-bait	1	28	not tested	-	-	28
bait 26	PRO_{35S}:GS-bait	1	1	PRO_{GOS2}:bait-GSgr	1	1	2
bait 27	PRO_{35S}:GS-bait	1	2	PRO_{GOS2}:bait-GSgr	1	3	5
bait 28	PRO_{35S}:GS-bait	1	0	PRO_{GOS2}:bait-GSgr	1	6	6
bait 29	not tested	-	-	PRO_{GOS2}:bait-GSgr	1	0	0
bait 30	not tested	-	-	PRO_{GOS2}:bait-GS	1	3	3
bait 31	not tested	-	-	PRO_{GOS2}:bait-GS	1	2	2
bait 32	PRO_{GOS2}:GSgr-bait	1	-	PRO_{GOS2}:bait-GS	1	6	6
bait 33	PRO_{GOS2}:GSgr-bait	0	0	not tested	-	-	0

Overview of the results from screening of 10 baits in T0 shoots. Each bait protein was tagged in an N- and C-terminal fusion construct, but not all construct were effectively tested however. The construct for the TAP fusions is specified for each bait in the 'construct' column. In the 'bait' column, the number of times the bait was retrieved is given. The number of interactors shown reflect the interactors that found. As only one purification was performed per construct, these were not necessarily reproducibly found, hence the asterisk.

Figure S4. Overview of high-throughput screening in T1 seedlings

	N-terminal fusion construct			C-terminal fusion construct			
bait n°	**construct**	**bait**	**interactors**	**construct**	**bait**	**interactors**	**total interactors**
bait 3	PRO_{35S}:GS-bait	2	18	PRO_{GOS2}:bait-GSgr	2	6	18
bait 4	PRO_{35S}:GS-bait	2	4	PRO_{GOS2}:bait-GSgr	2	3	4
bait 7	PRO_{35S}:GS-bait	2	31	PRO_{GOS2}:bait-GSgr	2	16	32
bait 14	PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	7	7
bait 22	PRO_{GOS2}:GSgr-bait	0	0	PRO_{35S}:bait-GS	0	0	0

Overview of the results from screening of 5 baits in T1 seedlings. Each bait protein was tagged in an N- and C-terminal fusion construct. The construct for these fusions is specified for each bait in the 'construct' column. In the 'bait' column, the number of times the bait was retrieved is given. We only considered baits as retrieved for a certain construct if these were found in both purifications. The number of interactors shown reflect the interactors that were reproducibly found, i.e. in at least two independent purifications.

Figure S5. Overview of the protein kinases screened in our TAP-MS workflow.

	N-terminal fusion construct				C-terminal fusion construct			
bait n°	platform	construct	bait	interactors	construct	bait	interactors	total interactors
bait 1	callus cells	PRO_{35S}:GS-bait	2	3	PRO_{GOS2}:bait-GSgr	2	0	3
bait 4		PRO_{35S}:GS-bait	2	5	PRO_{GOS2}:bait-GSgr	2	4	5
bait 8		PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	2	2
bait 25		PRO_{GOS2}:GSgr-bait	2	0	PRO_{35S}:bait-GS	2	3	3
bait 4	T0 shoots	PRO_{35S}:GS-bait	1	2	not tested	-	-	2
bait 4	T1 seedlings	PRO_{35S}:GS-bait	2	4	PRO_{GOS2}:bait-GSgr	2	3	4

Illustration of the performance of our platform when screening rice kinases as bait. Each bait protein was tagged in an N- and C-terminal fusion construct. The construct for these fusions is specified for each bait in the 'construct' column. In the 'bait' column, the number of times the bait was retrieved is given. We only considered baits as retrieved for a certain construct if these were found in both purifications. The number of interactors shown reflect the interactors that were reproducibly found, i.e. in at least two independent purifications.

Supplementary background list and mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP).

The established background list based on re-occurrence of a protein over independent experiments from different bait groups can be found under filename SI_Chapter_5_background_list.xlsx. The mass spec files for APC10 and CDKD interaction data described in this chapter can be found in SI_Chapter_6_1_PRO-APC10_MS_data.xlsx and SI_Chapter_6_1_PRO-CDKD_MS_data.xlsx respectively.

Author contribution

The PhD candidate was in charge for generation of the TAP constructs, the transformations, the maintenance and upscaling of callus cell lines and plant material, and analysis of the mass spectrometry data. He also wrote the chapter. For cloning and transformation, CropDesign employees helped sustaining the platform. TAP purifications and mass spectrometry measurements were mainly done by employees of VIB.

References

1. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

2. Hubner NC, Mann M: **Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC).** *Methods* 2011, **53**:453–459.

3. Hutchins JR a, Toyoda Y, Hegemann B, Poser I, Hériché J-K, Sykora MM, Augsburg M, Hudecz O, Buschhorn B a, Bulkescher J, et al.: **Systematic analysis of human protein complexes identifies chromosome segregation proteins.** *Science* 2010, **328**:593–599.

4. Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, et al.: **An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes.** *Nat. Protoc.* 2015, **10**:169–187.

5. Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K: **Dissecting plant genomes with the PLAZA comparative genomics platform.** *Plant Physiol.* 2012, **158**:590–600.

6. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S: **AmiGO: online access to ontology and annotation data.** *Bioinformatics* 2009, **25**:288–289.

7. Keilhauer EC, Hein MY, Mann M: **Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS).** *Mol. Cell. Proteomics* 2015, **14**:120–135.

8. Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neirynck S, Buffel Y, et al.: **Targeted interactomics reveals a complex core cell cycle machinery in Arabidopsis thaliana.** *Mol. Syst. Biol.* 2010, **6**:397.

9. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

10. Bürckstümmer T, Bennett KL, Preradovic A, Schütze G, Hantschel O, Superti-Furga G, Bauch A: **An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells.** *Nat. Methods* 2006, **3**:1013–1019.

11. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

12. Nelissen H, Eeckhout D, Demuynck K, Persiau G, Walton A, van Bel M, Vervoort M, Candaele J, De Block J, Aesaert S, et al.: **Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf.** Plant Cell 2015, **27**:1605–1619.

13. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang C, Iwamoto M, Abe T, et al.: **Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics.** *Plant Cell Physiol.* 2013, **54**:e6.

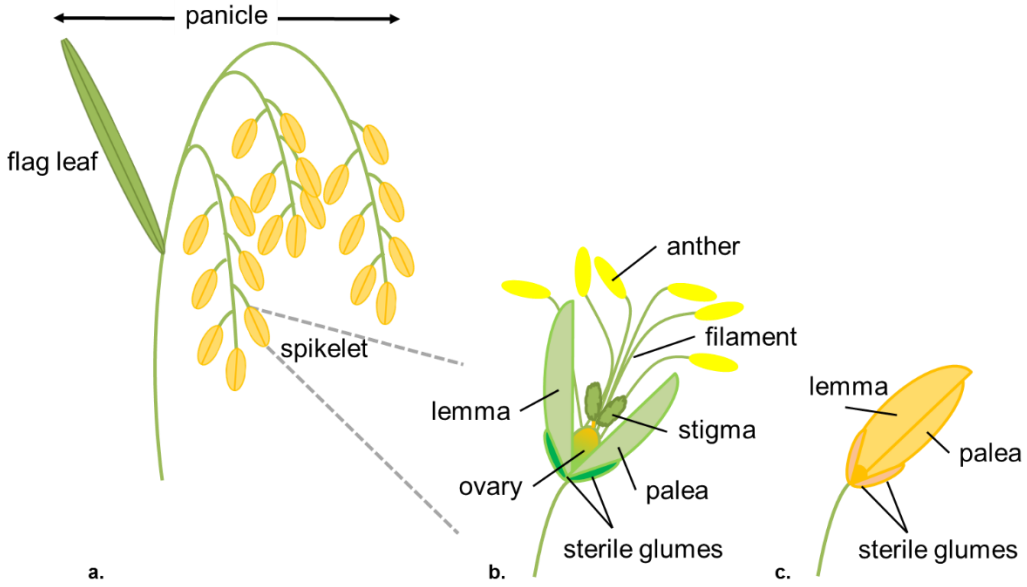
14. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al.: **Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data.** *Rice (N. Y).* 2013, **6**:4.

6.2 Ta_HLH, a lead gene linked to cell elongation and seed size

Importance of grain size for yield

Grain size is an important yield component. Seed production in rice is determined by inflorescence and grain development. The rice inflorescence or 'panicle' is the terminal component of the rice tiller (the 'rice stems'). A panicle bears rice spikelets, and it's the spikelets that develop into grains. The spikelet is the basic unit of the panicle and could be considered as the rice 'flower'. The floral organs are inside the lemma and the palea. The lemma is a larger protective glume covering the floral organs, the smaller one is referred to as the palea. A schematic overview of the rice inflorescence is shown in Figure 1. Once (self-) pollinated, the ovary ripens into a rice grain consisting of embryo and endosperm, with the lemma and palea firmly attached to it. This way they provide mechanical protection. On the other hand, the sizes of lemma and palea physically restrict the size of the grain (i.e., grain length, width, and thickness) [1]. Therefore, with ideal grain filling, the size of the lemma and palea determines the final grain weight.

Figure 1. Overview of the rice inflorescence



a. A rice panicle, bearing multiple spikelets. **b.** Floral organs in the spikelet during flowering. **c.** Rice grain, once (self-) pollinated, the ovary ripens into a rice grain consisting of embryo and endosperm, with the lemma and palea firmly attached to it.

Recent studies have identified some underlying QTLs (Quantitative Trait Loci) for grain weight positively influencing crop yields [2–12]. However, the current understanding of the mechanisms of grain weight regulation remains fragmentary, and the precise mechanism by which each of the yield-enhancing components function is unknown.

Results from TraitMill

We screened a *Triticum aestivum* basic helix-loop-helix (bHLH) gene, Ta_HLH, in TraitMill under control of the medium constitutive PRO_{GOS2}. Although considered a

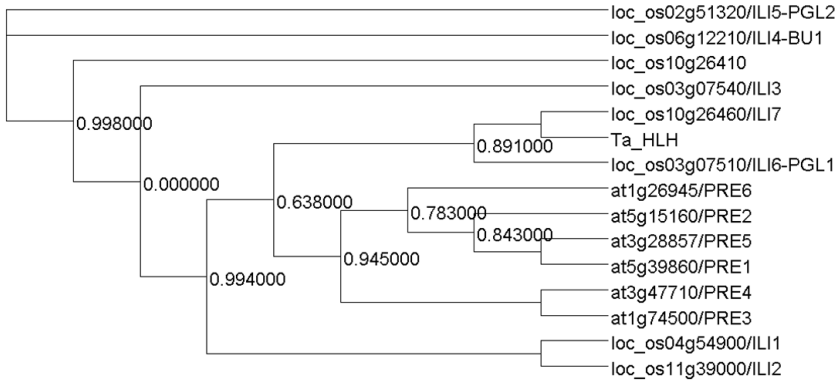
bHLH, the protein lacks the basic domain necessary for DNA binding. Three independent phenotypic yield screens showed a clear positive effect on seed size. Indeed, the Thousand Kernel Weight (TKW) parameter was significantly increased in the transgenes compared to the null segregants for all three screens. Concomitant however, there was a negative effect on the number of panicles and thus flowers and on the fill rate. Fill rate is an indication of filling of the seeds and is defined as the proportion of number of filled seeds over the number or florets. Overall the overexpression of the HLH thus resulted in a neutral to negative yield effect.

Phylogenetic analysis

Based on sequence similarity, Ta_HLH is closest related to rice loc_os10g26460 (Figure 2a). This gene was previously denoted as *ILI7*, part of a seven member gene family based on similarity to *INCREASED LEAF INCLINATION 1* (*ILI1*) [13]. The Arabidopsis homologs to the ILI family members were previously denoted PACLUBUTRAZOL RESISTANCE (PRE). Arabidopsis PRE and rice ILI families appeared to have evolved after separation of monocots and dicots suggesting functional specification related to specific developmental processes [13].

Figure 2. Overview of the homologs of TA_HLH in Arabidopsis and rice.

a.



b.

	PRE/ILI member	accession		also known as	
Arabidopsis	*PRE1*	at5g39860	*BNQ1*	bHLH136	
	PRE2	at5g15160	*BNQ2*	bHLH134	
	PRE3	at1g74500	*ATBS1*	bHLH135	*TMO7*
	PRE4	at3g47710	*BNQ3*	bHLH161	
	PRE5	at3g28857			
	PRE6	at1g26945	*KDR*		
Rice	*ILI1*	loc_os04g54900			
	ILI2	loc_os11g39000			
	ILI3	loc_os03g07540			
	ILI4	loc_os06g12210	*BU1*		
	ILI5	loc_os02g51320	*PGL2*		
	ILI6	loc_os03g07510	*PGL1*		
	ILI7	loc_os10g26460			
		loc_os10g26410			

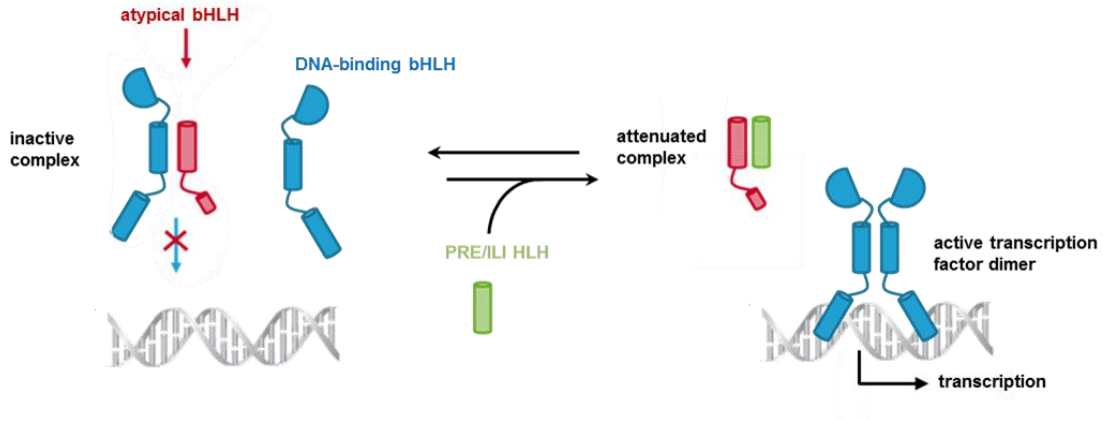
a. Phylogenetic tree illustrating the relation of TA_HLH with the rice ILI and Arabidopsis PRE families. **b.** Overview of the PRE and ILI family members in Arabidopsis and rice respectively, and their aliases. BNQ: BANQUO; ATBS1: ACTIVATION-TAGGED SUPPRESSOR OF BRI1-103; TMO7: TARGET OF MONOPTEROS 7; KDR: KIDARI; BU1: BR UPREGULATED 1; PGL: PROMOTER OF GRAIN LENGTH.

The PRE/ILI family of proteins represent 'atypical' bHLH's as they do not contain a basic domain and consequently are presumed not to possess DNA binding activity [13] (Figure S1). Instead, they regulate the activity of other bHLH's by heterodimerisation [13]. By doing so, they perturb the formation of functional protein dimers (or multimers) by forming non-functional protein complexes with their targets, which they regulate in a dominant negative manner. These proteins were referred to as 'microProteins' because their actions are analogous to microRNAs which are negative regulators of mRNAs [14]. Similar action of microProteins also evolved in the animal kingdom, with the Id- and Id2-proteins regulating cell-fate choices [15–17] and circadian clock [18] respectively. The concept of microProteins was further extended from atypical bHLH proteins to other types of proteins which retained a dimerization domain, but lacked a functional domain [14].

Arabidopsis literature

In Arabidopsis, at least three of the PRE family members act in tripartite modules, positively regulating cell elongation [19–22] (Figure 2). According to this mechanism, the PRE proteins attenuate a negative regulator by competitively forming non-functional heterodimers, releasing a transcription factor from the transcriptionally inactive complex. The activities of this tri-antagonistic module occur downstream of several hormonal and environmental stimuli, suggesting that these signals converge on this cascade.

Figure 2. Illustration of the tripartite module Arabidopsis PRE proteins are participating in.



Genuine bHLH proteins can bind DNA and activate downstream target genes. An atypical bHLH protein suppresses bHLH activity by forming non-DNA-binding complexes with the bHLH monomers. The plant PRE/ILI HLH protein sequesters the atypical bHLH by forming an attenuated complex, and shifts the equilibrium to bHLHs allowing them to form active homodimers. Plant PRE/ILI family proteins regulate cell elongation and expansion in response to various hormonal and/or environmental factors.

PRE1 for example binds the negative regulator INCREASED LEAF INCLINATION1 BINDING BHLH (IBH1) [13]. *PRE1* expression is activated by brassinosteroids (BRs) and gibberellic acid (GA) and repressed by light [23,24], whereas IBH1

function is negatively regulated by BRs [13]. IBH1, a HLH-type protein, lacks a basic domain required for binding DNA and acts as a transcriptional repressor. It interacts with regular bHLH transcription factors and prevents their homodimerization required for binding DNA and activating gene expression [19,20] Several bHLH proteins, including HOMOLOG OF BEE2 INTERACTING WITH IBH1 (HBI1), ACTIVATOR FOR CELL ELONGATION 1 (ACE1), ACE2, ACE3 and CRYPTOCHROME INTERACTING BHLH 1 (CIB1) are thus inhibited by IBH1 through heterodimerization [19–21]. HBI1, ACEs and CIB1 positively regulate cell elongation in response to BRs, GAs, temperature and light signalling [19–21]. In the proposed triantagonistic model HBI1, ACEs and CIB1 promote expression of cell elongation stimulating factors (e.g. expansins) and are negatively regulated through dimerization with IBH1 [19,20]. IBH1 itself can be sequestered through dimerization with PRE1, releasing ACEs and HBIs for homodimerization and promotion of cell elongation.

A similar triantagonistic model involves the PRE family member *ACTIVATION-TAGGED BRI1-301 SUPPRESSOR1 (ATBS1,* 80.65% coverage, 85.71% similarity to TA_HLH) and the putative non-DNA binding HLH proteins ATBS1-INTERACTING FACTOR1 (AIF1), AIF2, AIF3 and AIF4. *ATBS1* upregulation was shown to complement the bri1-301 phenotype (hence, *ACTIVATION-TAGGED BRI1-301 SUPPRESSOR1* or *ATBS1*), which is a weak mutant of the brassinosteroid receptor BRI1 (BRASSINOSTEROID INSENSITIVE 1). Therefore *ATBS1* potentially plays a role in BR signalling [25]. *ATBS1* was also shown to be a target of MONOPTEROS in root initiation and is potentially involved in auxin signalling [26]. AIFs negatively regulate BR signalling and cell elongation through heterodimerisation with DNA-binding bHLH transcription factors, and are antagonized by PRE1 and ATBS1 [21].

PRE-member proteins also regulate cell elongation in response to light signalling. Overexpression of *KIDARI* (*PRE6*) reduces light sensitivity and causes a long hypocotyl phenotype. Expression of *KIDARI* shows circadian oscillation, and KIDARI is proposed to mediate circadian regulation of light-responsive cell elongation by inhibiting LONG HYPOCOTYL IN FAR RED 1 (HFR1) [22,27]. HFR1 is a bHLH transcription factor that plays a role in plant photomorphogenesis by forming non-DNA binding heterodimers with PHYTOCHROME INTERACTING FACTORS (PIFs). Similarly, PRE1 antagonizes PHYTOCHROME RAPIDLY REGULATED (PAR), which inhibits PIF4 [28]. Hence both PRE1 and KIDARI are also in this case involved in a tripartite transcriptional regulation module. The PRE1 and KIDARI proteins attenuate PAR and HFR1 activity respectively by competitively forming non-functional heterodimers, causing liberation of PIFs from the transcriptionally inactive PAR-PIF and HFR1-PIF complexes.

Rice literature

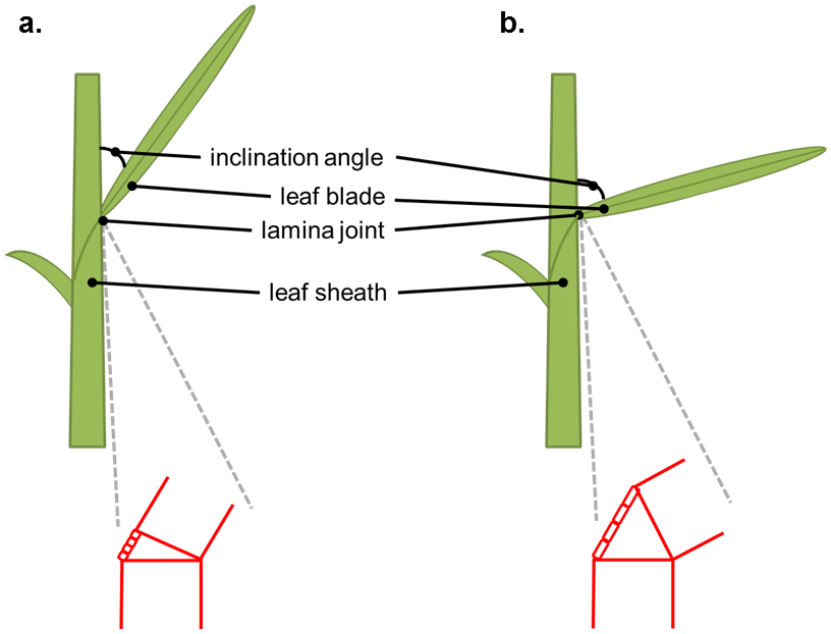
The PRE-like family seems evolutionary conserved in encoding transcriptional regulators that control cell elongation and expansion in specific organs. Tomato *STYLE2.1* for example controls cell elongation in the developing tomato style [29]. In rice, overexpression of PRE-homologs *INCREASED LEAF INCLINATION 1* (*ILI1*), *BRASSINOSTEROID UPREGULATED 1* (*BU1*), *PROMOTER OF GRAIN LENGTH 1* (*PGL1*) and *PGL2* resulted in an increased lamina inclination and enhanced grain

size phenotype [13,30,31,32]. The lamina joint is a border region between the leaf blade and sheath, and is an especially sensitive organ to brassinosteroids. It is severely bent after exposure to active BRs [33] (Figure 3). The increased lamina joint inclination is the result of greater cell expansion of adaxial cells relative to the adorsal cell in the joint region (Figure 3). Leaf angle is an important trait in cereal crops as it allows higher density planting and therefore can have a major impact on biomass and grain yield per hectare [34]. Indeed, rice plants with an erect leaf phenotype and no negative effect on seed size were estimated to lead to a ~30% increase in grain yield under high planting density [35]. Therefore, control of bending of the lamina joint via regulating expression of PRE-like proteins may lead to increased biomass and grain yield. In addition, overexpression of *BU1*, *PGL1* and *PGL2* display a valuable phenotype, namely enlarged grains. There is also convincing evidence for a role of BRs in plant seed production. Knock-out or downregulation of non-redundant BR biosynthetic genes and positive regulators of the BR signalling pathway in rice generally result in sterility and/or strongly reduced seed yield due to smaller and rounder seeds [35–41]. In contrast, rice transgenic plants overexpressing the transcription factors *BRASSINOSTEROID UPREGULATED 1*, a positive regulator of BR signalling have enlarged seeds [42]. Whereas the molecular mechanisms by which BRs affect seed size remain to be determined [43], overexpression of the *PRE*-like genes in rice led to larger grains due to increased cell elongation in the lemma and palea [31,32]. Further studies of downstream effects of these proteins may thus provide important insights for engineering the plant's architecture and increasing grain yield.

BU1 is a positive regulator of BR responses. It is a primary response gene that participates in BR-signalling pathways and controls bending of the lamina joint downstream of BR signalling components OsBRI1 (for *Oryza sativa* BRASSINOSTEROID INSENSITIVE 1) and OsD1/RGA1 (for Rice heterotrimeric G-protein α-subunit) [30].

ILI1 interacts with OsIBH1 to regulate rice leaf angle antagonistically. The brassinosteroid-regulated transcription factor BZR1 (BRASSINAZOLE-RESISTANT 1) directly binds to the promoter of these two genes to induce *ILI1* and represses *IBH1* expression [13]. In fact, the Arabidopsis representatives, *PRE1* and *AtIBH1* were identified through searching for orthologous genes of rice *ILI1* and *OsIBH1*. Analogous to their orthologs in Arabidopsis, rice ILI1 binds and inhibits OsIBH1 activity to regulate cell elongation. The conserved function of ILI1/PRE1 and IBH1 in rice and Arabidopsis suggests that there could be a rice ortholog of *HBI1* involved in the BR-signalling pathway to control leaf bending.

Figure 3. Effect of brassinosteroids on leaf inclination in rice.



This figure illustrates the effect of brassinosteroids on the inclination angle of the lamina joint. **a.** The untreated or wild type situation. **b.** The leaf bending phenotype after induction or upregulation of brassinosteroid signalling. The red drawings show the lamina joint, the linker between leaf blade and leaf sheath, mediating leaf bending. In plants where brassinosteroid signalling is induced or upregulated, adaxial cell elongation will be promoted, leading to plants that show bended leafs.

A closely similar rice protein to TA_HLH is PROMOTER OF GRAIN LENGTH, PGL1. PGL1 was found in an antagonistic pair of basic helix-loop-helix (bHLH) proteins determining rice grain length through regulation of cell elongation in the lemma and palea cells [31]. Consequently, overexpression of PGL1 in lemma/palea increased grain length and weight in transgenic rice. Similarly, grain-specific overexpression of *POSITIVE REGULATOR OF GRAIN LENGTH 2* (*PGL2*) increased grain length [32]. Both PGL1 and PGL2 counteract ANTAGONIST OF PGL1 (APG) through heterodimerisation. Silencing of APG produced the same phenotype as overexpression of PGL1, suggesting antagonistic roles for the two genes. While APG retains a basic domain and the necessary amino acids for DNA binding, actual transcriptional activation/repression was not tested.

So far, only two-component antagonistic pairs were uncovered in rice. It is not clear whether these take part in an evolutionary conserved tripartite module, similar to Arabidopsis. Further, rice PRE-homologs were only confirmed to function in the context of BR signalling. E.g. only Arabidopsis *PRE1* but not rice *ILI1* was induced by GA and auxin [13]. Whether GA and auxins regulate other members of the PRE/ILI family in rice remains to be tested.

Results

The aim of the experiment was in fact twofold. First we wanted to find interaction partners for the Ta_HLH lead gene. Second, we were seeking for further validation of our assay. Therefore part of the proteins we assayed overlap with baits that were previously screened using Arabidopsis cell suspension cells by the Functional Interactomics group at the 'Vlaams Instituut voor Biotechnologie' according to the protocol of Van Leene *et al.,* 2008 [44]. We expected that at least part of the interactors found in the suspension cells would overlap with homologous interacting proteins in rice. For these reasons, the Ta_HLH protein was assayed in both callus cells and whole seedlings. Although the latter might be less informative in relation to yield (as the yield effect was mainly related to seed size), it could help us assess our protocol in this plant tissue.

To identify interacting partners of TA_HLH, we created a TAP-tagged version of the protein for tandem affinity purification coupled to mass spectrometry (TAP-MS). We designed translational fusions with an N-terminal GSgreen tag expressed by PRO_{GOS2}, and with a C-terminal GS tag under control of PRO_{35S}. These constructs were introduced in rice callus and T1 seedlings. Expression of the TAP constructs in the transgenic tissues was verified at the protein level before performing purification experiments (Figure S2).

We executed in total nine purifications on the two types of material. From transgenic callus cells expressing the C-terminal GS-tagged bait, we performed two replicate experiments on 50mg total protein. Transgenic callus expressing the GSgreen-tagged bait was used for two replicate experiments on 25 mg and one experiment on 50mg total protein. In parallel, we screened the bait constructs in plant tissues. Transgenic seeds from 60 independent transgenic events were sown to grow 2 weeks old seedlings ('T1 seedlings') expressing the baits for duplicate purifications on both TAP fusions. We could further compare the results (Figure 3a) with interaction data from previous experiments using C- and N-terminal fusions of TA_HLH to the GS TAP tag expressed under the control of the 35S promoter in Arabidopsis cell cultures (Figure 3b).

Figure 3. Interactions identified from experiments in Arabidopsis and rice tissues.

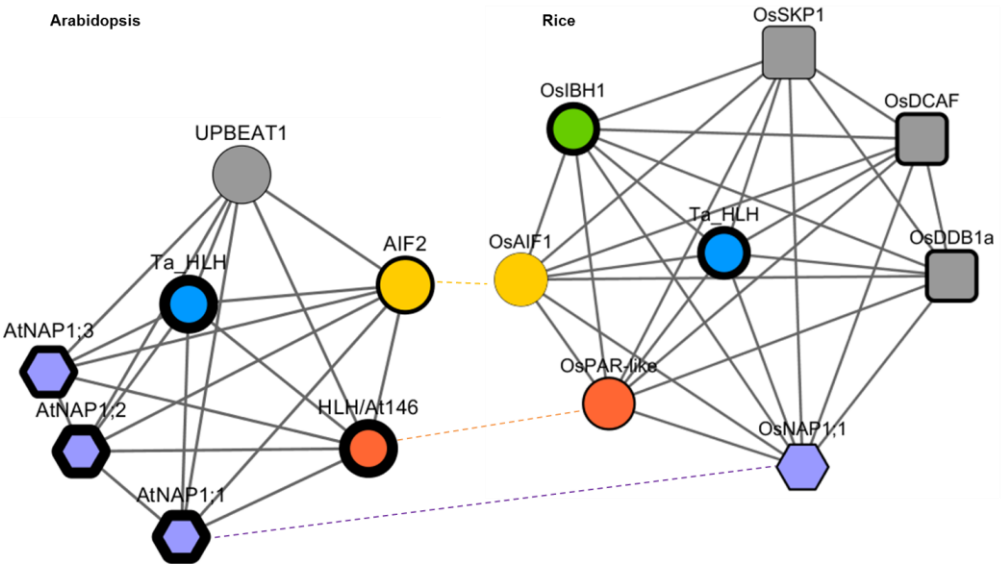
a.

	accession	description	symbol	PRO_{GOS2}-NGSgreen (2)	PRO_{35S}-CGS (2)	PRO_{GOS2}-NGSgreen* (1)	PRO_{GOS2}-NGSgreen (2)	PRO_{35S}-CGS (2)	total
bait	loc_os03g07510	PROMOTER OF GRAIN LENGTH1	PGL1	2	2	1	2	2	9
	loc_os10g26410	INCREASED LEAF INCLINATION 7	ILI7					2	2
transcriptional regulators	loc_os04g56500	ILI1-BINDING BHLH1	IBH1	2	2	1	2	1	8
	loc_os01g43950	PAR-like			2	1			3
	loc_os01g14110	AIF1-like				1			1
protein degradation	loc_os04g59460	DCAF/DWD PROTEIN	DCAF/DWD	2	2	1			5
	loc_os05g51480	DDB1A	DDB1A	2	2	1			5
	loc_os09g36830	ARABIDOPSIS SKP1-LIKE			2				2
DNA structure	loc_os06g05660	NUCLEOSOME ASSEMBLY PROTEIN 1;1	NAP1;1	2		1			3
	loc_os05g49200	ASPARTIC PROTEASE		2					2
	loc_os07g44440	1-CYS PEROXIREDOXIN			2				2

b.

				Arabidopsis cell suspensions		
	accession	description	symbol	PRO_{35S}-NGS (2)	PRO_{35S}-CGS (2)	total
bait	loc_os03g07510	PROMOTER OF GRAIN LENGTH1	PGL1	2	2	4
transcriptional regulators	at4g30180	HLH/At146		2	2	4
	at3g06590	ATBS1 Interacting Factor 2	AIF2	2		2
	at2g47270	UPBEAT1	UPB1		1	1
DNA structure	at4g26110	NUCLEOSOME ASSEMBLY PROTEIN1;1	NAP1;1	2	2	4
	at2g19480	NUCLEOSOME ASSEMBLY PROTEIN1;2	NAP1;2	2	2	4
	at5g56950	NUCLEOSOME ASSEMBLY PROTEIN1;3	NAP1;3	2	1	3

c.



a. Interaction table for proteins retrieved from TAP experiments from rice T0 callus, T0 shoots and T1 seedlings. For GSgreen constructs, purifications were performed on 25mg and 50mg (asterisk) of total protein from transgenic callus. **b.** TAP interaction data from Arabidopsis. The number of replicate experiments for a construct are shown between brackets. Proteins identified from the rice protein annotation database with peptides matching to the TA_HLH sequence are shown in blue. Proteins that were not reproducibly found are shown in grey. Proteins that were identified in experiments with more than two other unrelated bait groups, and normally would have been considered non-specific/false positive, are shown in red. **c.** Commonalities between Arabidopsis and rice interactions. The protein type is reflected in the shape of the nodes; transcriptional regulators are shown as ellipses, proteins involved in histone chaperoning as hexagons and proteins regulating protein degradation as rounded squares. The border width of the nodes reflects the amount of times the protein was found over the purifications. The node colours illustrate the protein family to which the protein belongs: yellow: AIF family of HLH proteins; orange: PAR-like family or HLH proteins; green: IBH1-like family of HLH proteins; blue: PRE-like family of HLH proteins; purple: NAP family; grey: orphan (UPBEAT1) or not checked.

Since TA_HLH is a wheat protein, its sequence is not annotated in the searched databases. Therefore it was only indirectly identified through its rice homologs PGL1 and ILI7. In most of the experiments the bait was characterised as PGL1. In two experiments from seedlings, it was also found as ILI7. As no peptides were identified that uniquely match to PGL1 or ILI7 protein sequences, the identifications most likely reflect detection of the bait. We cannot exclude however that also the respective PRE-like members were found. We therefore propose that these identifications should be looked at as possible isoforms whilst considering that mutual interactions between PRE family members were not observed previously.

Despite the evolutionary distance between Arabidopsis and rice, we still found reasonable overlap between the interaction datasets. Two families of HLH proteins had representatives in both datasets. Loc_os01g14110 and at3g06590 are both member of the PAR-like subfamily of bHLH's, or subfamily 21 as proposed in [45]. Other members of this subfamily are *PHYTOCHROME RAPIDLY REGULATED 1* (*PAR1*) and *PAR2*, which act as negative regulators of a variety of shade avoidance syndrome (SAS) responses. As such, PAR1 and PAR2 antagonise conventional bHLH transcription factors by forming heterodimers, and preventing their binding to DNA or other transcription factors [46,47] (Figure 4b). If the PAR-like interactors we identified act in a similar mechanism, this could hint towards the existence of triantagonistic regulation mechanisms in rice. In that case, TA_HLH would act as a positive regulator by counteracting the PAR-like negative regulators, and releasing conventional bHLH transcription factors to exert their function in activating gene expression. PRE1 and KIDARI, Arabidopsis orthologs of the bait protein, were already shown to sequester PAR1 in a module that controls skotomorphogenesis (seedling development in the dark) and shade avoidance [28].

Another family represented in both Arabidopsis and rice datasets is the ATBS1-Interacting Factor (AIF) family of HLH proteins. We found a rice ortholog of AIF1 and Arabidopsis AIF2 in callus and cell suspension cultures respectively. AIF1, for which we found the rice ortholog, differs from the other three Arabidopsis AIFs as it does not interact with the HLH PRE1 or the conventional bHLH ACTIVATOR FOR

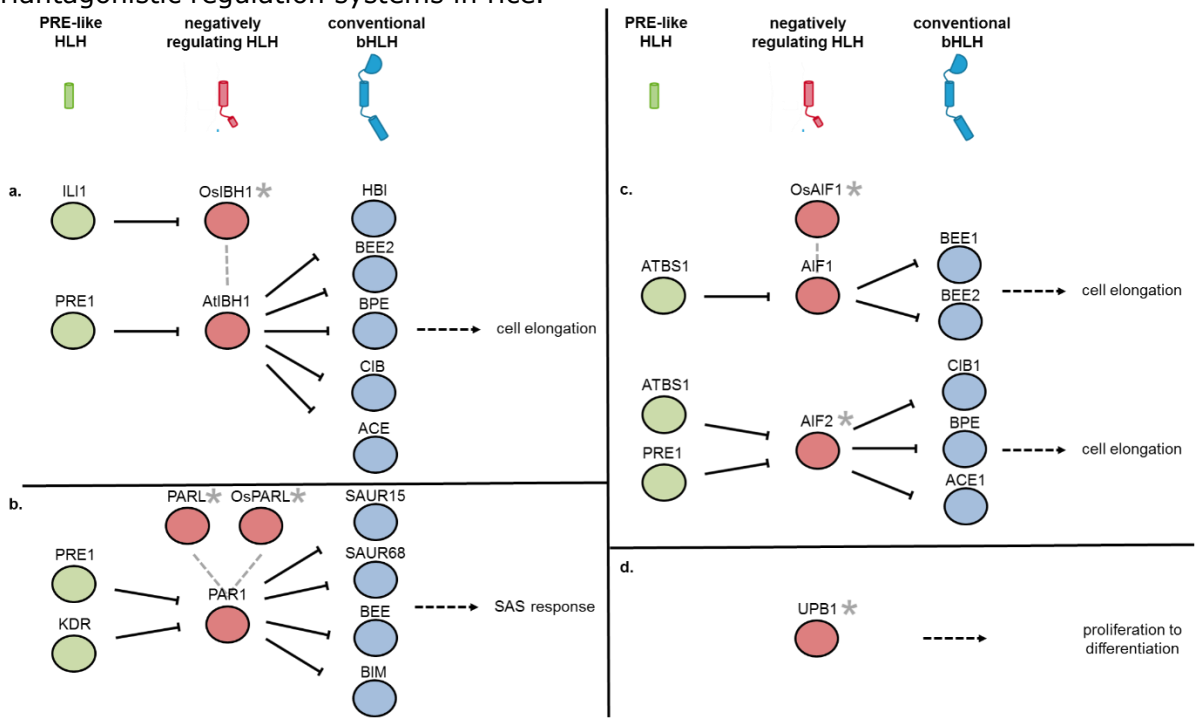
CELL ELONGATION1 (ACE1) [21]. It was also the only AIF member involved in BR signalling. AIF2 does interact with PRE1 in Y2H assay, and also potentially regulates DNA binding activities of BIG PETAL (BPE) and CHRYPTOCHROME INTERACTING BHLH 1 (CIB1) in addition to that of ACE transcription factors (Figure 4c).

AIFs were included in a larger family of atypical bHLH subfamily together with Arabidopsis *IBH1* and *UPB1* [19]. We found the OsIBH1 protein in rice purifications and the UPBEAT1 protein in Arabidopsis purifications. Similar as to what was found for the PAR-like interactors, the identification of the rice ortholog of IBH1 suggests that a trimodular regulation system is present in rice (Figure 4a). OsIBH1 and its Arabidopsis counterpart differ however in some aspects. Arabidopsis IBH is an integration point of different regulation mechanism involving BRs, GAs, light and temperature [23,24,48], whereas for OsIBH1 only involvement in BR signalling could be determined [13]. The UPBEAT1 (UPB1) HLH protein found only in Arabidopsis cell suspension cells was shown to control the balance between cellular proliferation and differentiation through transcriptional regulation [49] (Figure 4d). Strikingly, UPBEAT1 seemed to stimulate rather than inhibit elongation in contrast to IBH1, PARs or AIFs. UPB1 binds DNA and effectively activates transcription of a set of peroxidases that modulate the balance of reactive oxygen species (ROS) between the zones of cell proliferation and the zone of cell elongation where differentiation begins [49].

Apart from HLH proteins, we found another class of potential transcriptional regulators. Indeed, we co-purified NAP-(for NUCLEOSOME ASSEMBLY PROTEIN) domain containing proteins in both Arabidopsis and rice. NAPs are thought to act as histone chaperones, shuttling both core and linker histones from their site of synthesis in the cytoplasm to the nucleus. Therefore these proteins may be involved in regulating gene expression.

Apart from proteins involved in transcription regulation we retrieved the DDB1-CUL4 ASSOCIATED FACTOR (DCAF) together with UV-DAMAGED DNA BINDING PROTEIN (DDB1A) and S-PHASE KINASE ASSOCIATED PROTEIN 1 (SKP1) from callus. Human DCAF proteins have been reported to interact directly with DDB1 through the WDxR motif in their WD40 domain and function as substrate-recognition receptors for CULLIN4-based E3 ubiquitin ligases [50]. Also in Arabidopsis, the homolog of human DCAF1 was found to physically interact with DDB1 in Y2H assay [51].

Figure 4. Homologies with Arabidopsis components hint towards the existence of triantagonistic regulation systems in rice.



Proteins identified as potential interactors of TA_HLH (marked with an asterisk) all belong to the class of negatively regulating HLH proteins. Data from Arabidopsis hints towards participation of TA_HLH in trimodular regulation systems involving the identified interactors and conventional bHLH's as depicted in the figure. **a.**, **b.** and **c.** highlight the regulatory mechanisms where IBH1, PAR and AIF subfamily members of HLH proteins participate in respectively. **d.** UPBEAT1 seems to be an exceptional case, in which the HLH protein is able to bind DNA and activate transcription of peroxireductases in the switch from proliferation to differentiation in Arabidopsis roots. Grey dashed line marks homology between the proteins. SAS: Shade Avoidance Syndrome response; ILI1: INCREASED LEAF INCLINATION 1; IBH1: INCREASED LEAF INCLINATION1 BINDING BHLH; HBI: HOMOLOG OF BEE2 INTERACTING WITH IBH1; BEE: BR ENHANCED EXPRESSION ; BPE: BIG PETAL; CIB: CHRYPTOCHROME INTERACTING BHLH 1; PRE1: PACLUBUTRAZOL RESISTANT 1; KDR: KIDARI; PARL: PAR-LIKE; PAR1: PHYTOCROME RAPIDLY REGULATED 1; SAUR: SMALL AUXIN UPREGULATED; BIM: BES1-INTERACTING MYC-LIKE; ATBS1: ACTIVATION-TAGGED SUPPRESSOR OF BRI-1 301; AIF: ATBS1-INTERACTING FACTOR; ACE: ACTIVATOR OF CELL ELONGATION; UPB1: UPBEAT 1.

Conclusions

Rice grain is tightly enclosed by a hull composed of lemma, palea and two empty glumes. During the process of grain filling, endosperm cells expand and accumulate a massive amount of nutrients, mainly starch. Rice grain weight is largely determined by the endosperm size, which is linked to the constraints determined by lemma and palea size. Dozens of genes or QTL's involved in rice grain weight were isolated and characterized. In an ideal situation where the entire panicle is fully filled, grain weight is determined by grain size.

PGL1 and APG encode an antagonistic pair of bHLH proteins that interact to regulate rice grain length [31]. PGL1 and PGL2 redundantly suppress the function of APG to form elongated grains [32]. Phenotypic screening of an overexpression

construct of TA_HLH, a wheat homolog of PGL1 and PGL2, in our TraitMill platform confirmed a positive effect on seed size. This coincides however with a negative effect on the number of panicles and flowers resulting in an overall neutral yield effect. The potential of the PRE/ILI family in increasing seed yield is reflected in the recent identification of the GW6 (Grain Weight 6) QTL [12]. This quantitative trait locus encodes a GCN5-related N-acetyltransferases (GNAT)-like protein that harbours intrinsic histone acetyltransferase activity (OsIHAT1). The favourable allele of OsIHAT1 resulted in elevated expression, enhancing grain weight and yield by enlarging spikelet hulls through increased cell number and accelerated grain filling, and increasing global acetylation levels of histone H4. OslHAT1 localizes to the nucleus, where it likely functions through the regulation of transcription. Interestingly, transgenic overexpression of OslHAT1 resulted in activation of PGL2 expression [12]. Analysis of ILI1 and BU1, two other members of the family of HLH proteins to which PGL1 and PGL2 also belong, in rice hinted towards involvement of the BR pathway in affecting rice grain size. A series of mutants related to the synthesis and signalling pathway of BR such as *d61*, *brd1 (brassinosteroid dependent 1)* and *short grain1* (*sg1*) display shorter grain phenotype than their wild types [52–54] and confirm the potential involvement of BR in determining kernel phenotype.

To gain insight in potential pathways to increase seed size and make an assessment to avoid negatively affecting the panicle and flower number parameters, we screened for interaction partners of the TA_HLH gene. TAP experiments were performed both in Arabidopsis cell suspension cultures, rice callus cells and two weeks old rice seedlings. We retrieved different subfamilies of HLH-proteins from Arabidopsis as well as from rice experiments, and found that some could be traced back to matching subfamilies. As such, we identified representatives of the AIF and PAR families of HLH proteins in both species. In addition, we could retrieve UPBEAT1 in Arabidopsis, and OsIBH in rice. We found a representative of the AIF-subfamily and the PAR-subfamily of HLH proteins in both Arabidopsis and rice. Proteins from these families are considered negative regulators of transcription. They act through heterodimerization with conventional bHLH's and prevent these as such to form DNA-binding homodimers that can activate gene transcription. In that case, TA_HLH would be part of different tripartite modules regulating transcription as described in Arabidopsis. Such a triantagonistic model could also be in place for the OsIBH uniquely found in rice tissues. We further found UPBEAT1 uniquely in Arabidopsis cells. This is the only HLH protein in the dataset for which actual DNA binding and gene activation was observed [49].

Surprisingly, ANTAGONIST OF PGL1, a direct interaction partner of PGL1 in rice picked up by yeast two-hybrid was not retrieved in our experiments. Although the overlap between datasets from Y2H and TAP is only modest (10%, [55]), we don't expect that this is the reason of not detecting APG, since we were able to confirm other similar interactors that were established through Y2H. Instead we presume that the reason of failed detection is the seed specific expression pattern of APG [31]. From the transcriptional regulators we did co-purify, an AIF-homolog (100% coverage and 75.11% similarity with at3g06590) was tested in the TraitMill platform. PRO_{GOS2}-driven overexpression led to a significant negative effect for

seed size (TKW) compared to the null segregants. This opposite phenotype confirms that TA_HLH is antagonising the AIF-like regulators, at least in regulating the size of the seeds.

We were not able to identify the bHLH transcription factors that are downstream of the tripartite regulation modules and are responsible for transcriptional activation. Possibly binding of each (b)HLH component within the module is mutually exclusive. Therefore, to gain a comprehensive view on which transcription factors are most downstream of the module where TA_HLH takes part in, reciprocal TAP purifications using the IBH1, UPB1, AIFs and PARs as bait could be proposed. Further, the reasonably high number of different types of HLH-binding partners found, suggests that constitutive overexpression of TA_HLH may cause concomitant negative effects on the other yield parameters resulting in a generally neutral yield phenotype. PRE/ILI family members were indeed shown to be able to act redundantly in their function, and aberrant overexpression of an additional family member might result in pleiotropic effects. For example, UPBEAT1 attenuation in the roots by TA_HLH might explain the negative root phenotypes for 2 yield experiments, since UPB1 is responsible for control over expression of peroxidases that modulate the balance of reactive oxygen species (ROS) between the zones of cell proliferation and the zone of cell elongation where differentiation begins. Disruption of UPB1 activity alters this ROS balance, leading to a delay in the onset of differentiation.

Specifically from callus tissues, we co-purified a DCAF-DDB1 module and SKP1 with the TA_HLH bait. These are known to be part of substrate targeting units within distinct CULLIN-RING ubiquitin E3 ligases (CRL's); respectively the CUL4A RING (CRL4A) and SKP1-CULLIN1-F-box (SCF) complexes. In the former module, the DCAF protein is responsible for substrate recognition whereas the DDB1 acts as adapter protein, providing the link with CULLIN4. The C-terminal domain of CULLIN then binds to its RING partner, which recruits ubiquitin-loaded E2 enzymes for catalysis. The CULLIN nor the RING-box protein partner were retrieved in our purifications. SKP1 acts as the adapter protein for an alternative substrate targeting unit, an F-box protein within the SCF complex, linking it to CULLIN1. We failed to detect the F-box recognition unit in this complex. We also do not know of any previous interaction between the SKP1 adapter and DCAF forming a substrate targeting unit. At this point, it is unclear whether TA_HLH or PRE/ILI family members in general are recruiting this E3 ubiquitin ligase complex to target the negative regulator HLH for degradation, or if it is the PRE/ILI members themselves that are the targets for proteolysis. But apart from this, this illustrates that TAP can provide also insights in protein complex constitutions other than the initial complex started from, providing links to the surrounding protein machineries.

We conclude that tandem affinity purification experiments on the TA_HLH lead, a PRE/ILI member within the bHLH family resulted in the successful identification of a set of interactors. This dataset helped us in gaining insight in the biological functioning of the bait, and gave hints to potential pleiotropic participation in complexes containing different HLH family members. This could explain the negative phenotypes observed with the gain in kernel weight.

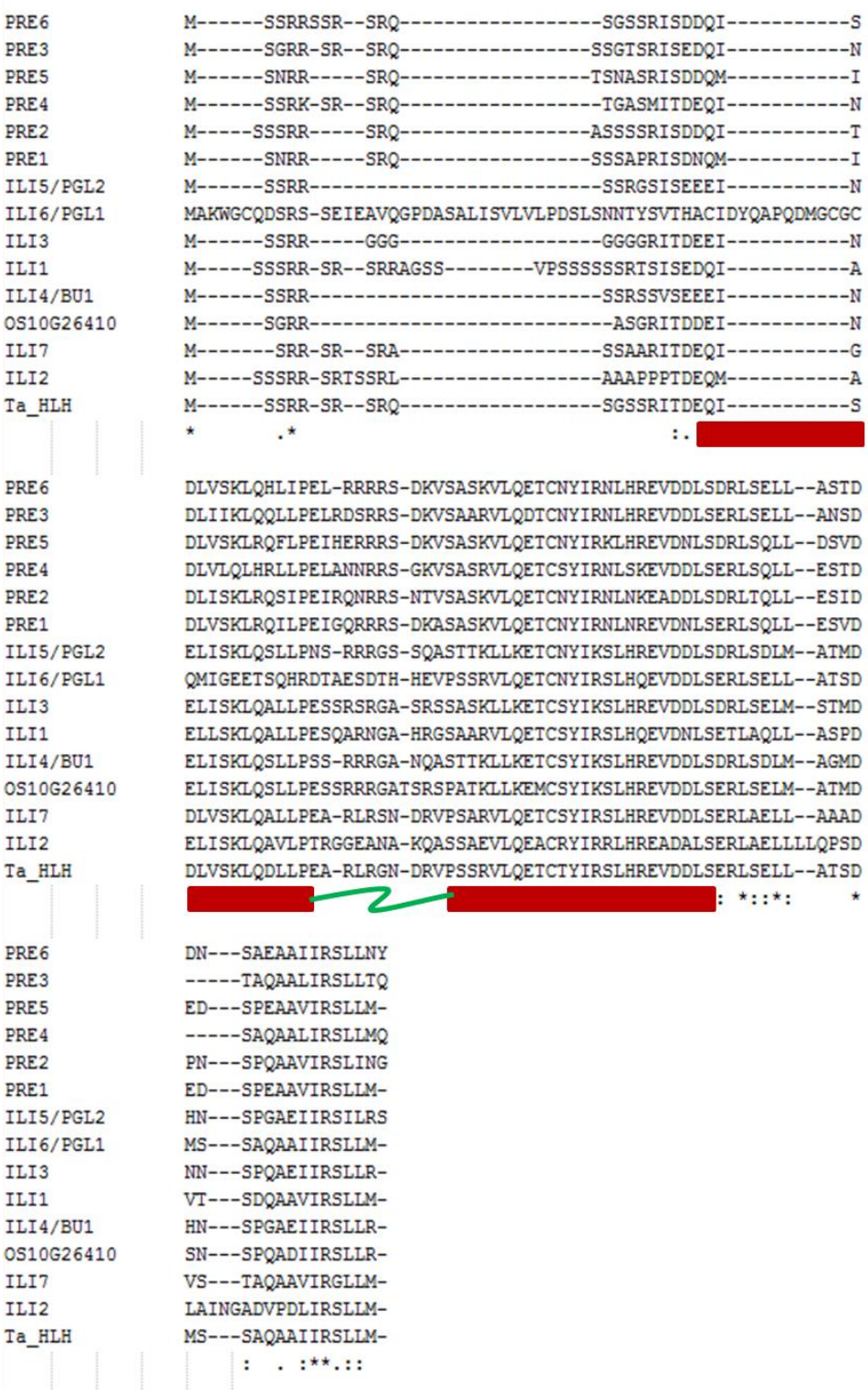
Materials & Methods

Phylogenetic analysis of the PRE/ILI family of HLH proteins
The protein sequences of the Arabidopsis PRE and rice ILI proteins used in the homology analyses were retrieved from TAIR and MSU databases respectively. The sequences were aligned with MAFFT [56] by using the L-INS-I option. This iterative refinement search sustains global multiple sequence alignments and integrates strong local alignments to enrich for alignment of functional domains. The Maximum Likelihood trees were calculated with PhyML (Version 3.0) [57]. The three was visualised using Dendroscope (Version 3.2.8) [58].

Screening of the TA_HLH bait in the rice TAP platforms
Screening of the bait in callus, T0 shoots and T1 seedlings was done as described in the materials and methods in chapter 6.1.

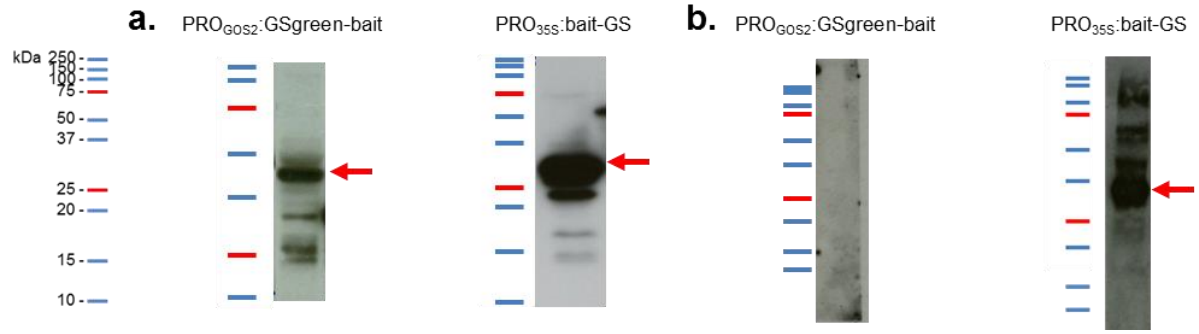
Supplementary information

Figure S1. The PRE/ILI-like proteins are HLH proteins.



MAFFT multiple amino acid sequence alignment of ILI and PRE family proteins and Ta_HLH. The red lines indicate the helix regions and the green line indicates the loop region.

Figure S2. Expression analysis of the bait protein.



Protein extracts of the transgenic lines were analyzed by immunoblotting with antibodies against the GS tag or GSgreen tag to verify protein levels of the TA_HLH bait. **a.** Bait expression levels from callus extracts. **b.** Expression levels from T1 seedling extracts. Molecular marker values are provided in kilo Dalton (kDa) in the left panel. The molecular weight of the tagged proteins is 45.6 kDa for the GSgreen-tagged bait and 38.5 kDa for the GS-tagged bait.

Figure S3. Primers used in this chapter.

primer n°	name	primer sequence
prm27139	attB1 kozak Ta_HLH	ggggacaagtttgtacaaaaaagcaggctccaccatgtcgagccgtaggtcaaggtc
prm27155	attB2 Ta_HLH no stop	ggggaccactttgtacaagaaagctgggtcggcagtgttgttgggttcag
prm27171	attB2r no start Ta_HLH	ggggaccactttgtacaagaaagctgggtggggacagctttcttgtacaaagtggccatgtcgagccgtagg
prm27187	attB3 Ta_HLH stop	ggggacaactttgtataataaagttggctacatcagcaagctgc

Supplementary mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP)

The mass spec files for TA_HLH interaction data described in this chapter can be found in SI_Chapter_6_2_TA_HLH_MS_data.xlsx.

Author contribution

The PhD candidate was in charge of generating the TAP constructs, the transformations, maintenance and upscaling of callus cell lines and plant material, analysis of the mass spectrometry data. He also wrote the manuscript. TAP purifications and mass spectrometry measurements were mainly done by employees of VIB.

References

1. Yoshida S: **Physiological Aspects of Grain Yield** *Annu. Rev. Plant Physiol.* 1972, **23**:437–464.

2. Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q: **Linking differential domain functions of the GS3 protein to natural variation of grain size in rice.** *Proc. Natl. Acad. Sci. U. S. A.* 2010, **107**:19579–19584.

3. Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X: **Natural variation at the DEP1 locus enhances grain yield in rice.** *Nat. Genet.* 2009, **41**:494–497.

4. Qi P, Lin Y-S, Song X-J, Shen J-B, Huang W, Shan J-X, Zhu M-Z, Jiang L, Gao J-P, Lin H-X: **The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3.** *Cell Res.* 2012, **22**:1666–1680.

5. Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, et al.: **Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:21534–21539.

6. Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X: **A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase.** *Nat. Genet.* 2007, **39**:623–630.

7. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M: **Deletion in a gene associated with grain size increased yields during rice domestication.** *Nat. Genet.* 2008, **40**:1023–1028.

8. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X, et al.: **Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight.** *Cell Res.* 2008, **18**:1199–1209.

9. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, et al.: **Natural variation in GS5 plays an important role in regulating grain size and yield in rice.** *Nat. Genet.* 2011, **43**:1266–1269.

10. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, et al.: **Control of grain size, shape and quality by OsSPL16 in rice.** *Nat. Genet.* 2012, **44**:950–954.

11. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B-I, Onishi A, et al.: **Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield.** *Nat. Genet.* 2013, **45**:707–711.

12. Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, Segami S, Miura K, Ogawa D, Kamura T, et al.: **Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice.** *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:76–81.

13. Zhang L-Y, Bai M-Y, Wu J, Zhu J-Y, Wang H, Zhang Z, Wang W, Sun Y, Zhao J, Sun X, et al.: **Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and Arabidopsis.** *Plant Cell* 2009, **21**:3767–3780.

14. Staudt A-C, Wenkel S: **Regulation of protein function by 'microProteins'.** *EMBO Rep.* 2011, **12**:35–42.

15. Atherton GT, Travers H, Deed R, Norton JD: **Regulation of cell differentiation in C2C12 myoblasts by the Id3 helix-loop-helix protein.** *Cell Growth Differ.* 1996, **7**:1059–1066.

16. Kondo T, Raff M: **The Id4 HLH protein and the timing of oligodendrocyte differentiation.** *EMBO J.* 2000, **19**:1998–2007.

17. Hacker C, Kirsch RD, Ju X-S, Hieronymus T, Gust TC, Kuhl C, Jorgas T, Kurz SM, Rose-John S, Yokota Y, et al.: **Transcriptional profiling identifies Id2 function in dendritic cell development.** *Nat. Immunol.* 2003, **4**:380–386.

18. Duffield GE, Watson NP, Mantani A, Peirson SN, Robles-Murguia M, Loros JJ, Israel MA, Dunlap JC: **A role for Id2 in regulating photic entrainment of the mammalian circadian system.** *Curr. Biol.* 2009, **19**:297–304.

19. Bai M-Y, Fan M, Oh E, Wang Z-Y: **A triple helix-loop-helix/basic helix-loop-helix cascade controls cell elongation downstream of multiple hormonal and environmental signaling pathways in Arabidopsis.** *Plant Cell* 2012, **24**:4917–4929.

20. Ikeda M, Fujiwara S, Mitsuda N, Ohme-Takagi M: **A triantagonistic basic helix-loop-helix system regulates cell elongation in Arabidopsis.** *Plant Cell* 2012, **24**:4483–4497.

21. Ikeda M, Mitsuda N, Ohme-Takagi M: **ATBS1 INTERACTING FACTORs negatively regulate Arabidopsis cell elongation in the triantagonistic bHLH system** *Plant Signal. Behav.* 2014, **8**:e23448.

22. Hong S-Y, Seo PJ, Ryu JY, Cho S-H, Woo J-C, Park C-M: **A competitive peptide inhibitor KIDARI negatively regulates HFR1 by forming nonfunctional heterodimers in Arabidopsis photomorphogenesis.** *Mol. Cells* 2013, **35**:25–31.

23. Oh E, Zhu J-Y, Wang Z-Y: **Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses.** *Nat. Cell Biol.* 2012, **14**:802–809.

24. Bai M-Y, Shang J-X, Oh E, Fan M, Bai Y, Zentella R, Sun T-P, Wang Z-Y: **Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in Arabidopsis.** *Nat. Cell Biol.* 2012, **14**:810–817.

25. Kang B, Wang H, Nam KH, Li J, Li J: **Activation-tagged suppressors of a weak brassinosteroid receptor mutant.** *Mol. Plant* 2010, **3**:260–268.

26. Schlereth A, Möller B, Liu W, Kientz M, Flipse J, Rademacher EH, Schmid M, Jürgens G, Weijers D: **MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor.** *Nature* 2010, **7290**:913-916.

27. Hyun Y, Lee I: **KIDARI, encoding a non-DNA Binding bHLH protein, represses light signal transduction in Arabidopsis thaliana.** *Plant Mol. Biol.* 2006, **61**:283–296.

28. Hao Y, Oh E, Choi G, Liang Z, Wang Z-Y: **Interactions between HLH and bHLH factors modulate light-regulated plant development.** *Mol. Plant* 2012, **5**:688–697.

29. Chen K-Y, Cong B, Wing R, Vrebalov J, Tanksley SD: **Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes.** *Science* 2007, **318**:643–645.

30. Tanaka A, Nakagawa H, Tomita C, Shimatani Z, Ohtake M, Nomura T, Jiang C-J, Dubouzet JG, Kikuchi S, Sekimoto H, et al.: **BRASSINOSTEROID UPREGULATED1, encoding a helix-loop-helix protein, is a novel gene involved in brassinosteroid signaling and controls bending of the lamina joint in rice.** *Plant Physiol.* 2009, **151**:669–680.

31. Heang D, Sassa H: **Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice.** *PLoS One* 2012, **7**:e31325.

32. Heang D, Sassa H: **An atypical bHLH protein encoded by POSITIVE REGULATOR OF GRAIN LENGTH 2 is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein APG.** *Breed. Sci.* 2012, **62**:133–141.

33. Wada K, Marumo S, Ikekawa N, Morisaki M, Mori K: **Brassinolide and Homobrassinolide Promotion of Lamina Inclination of Rice Seedlings** *Plant Cell Physiol.* 1981, **22**:323–325.

34. Sakamoto T, Morinaka Y, Ohnishi T, Sunohara H, Fujioka S, Ueguchi-Tanaka M, Mizutani M, Sakata K, Takatsuto S, Yoshida S, et al.: **Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice.** *Nat. Biotechnol.* 2006, **24**:105–109.

35. Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M: **Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice.** *Plant Physiol.* 2006, **141**:924–931.

36. Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A: **Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the alpha-subunit of GTP-binding protein.** *Proc. Natl. Acad. Sci. U. S. A.* 1999, **96**:10284–10289.

37. Mori M, Nomura T, Ooka H, Ishizaka M, Yokota T, Sugimoto K, Okabe K, Kajiwara H, Satoh K, Yamamoto K, et al.: **Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis.** *Plant Physiol.* 2002, **130**:1152–1161.

38. Hong Z, Ueguchi-Tanaka M, Umemura K, Uozu S, Fujioka S, Takatsuto S, Yoshida S, Ashikari M, Kitano H, Matsuoka M: **A rice brassinosteroid-deficient mutant, ebisu dwarf (d2), is caused by a loss of function of a new member of cytochrome P450.** *Plant Cell* 2003, **15**:2900–2910.

39. Hong Z, Ueguchi-Tanaka M, Fujioka S, Takatsuto S, Yoshida S, Hasegawa Y, Ashikari M, Kitano H, Matsuoka M: **The Rice brassinosteroid-deficient dwarf2 mutant, defective in the rice homolog of Arabidopsis DIMINUTO/DWARF1, is rescued by the endogenously accumulated alternative bioactive brassinosteroid, dolichosterone.** *Plant Cell* 2005, **17**:2243–2254.

40. Tanabe S, Ashikari M, Fujioka S, Takatsuto S, Yoshida S, Yano M, Yoshimura A, Kitano H, Matsuoka M, Fujisawa Y, et al.: **A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length.** *Plant Cell* 2005, **17**:776–790.

41. Oki K, Inaba N, Kitagawa K, Fujioka S, Kitano H, Fujisawa Y, Kato H, Iwasaki Y: **Function of the alpha subunit of rice heterotrimeric G protein in brassinosteroid signaling.** *Plant Cell Physiol*. 2009, **50**:161–172.

42. Tanaka A, Nakagawa H, Tomita C, Shimatani Z, Ohtake M, Nomura T, Jiang C-J, Dubouzet JG, Kikuchi S, Sekimoto H, et al.: **BRASSINOSTEROID UPREGULATED1, encoding a helix-loop-helix protein, is a novel gene involved in brassinosteroid signaling and controls bending of the lamina joint in rice.** *Plant Physiol*. 2009, **151**:669–680.

43. Vriet C, Russinova E, Reuzeau C: **Boosting crop yields with plant steroids.** *Plant Cell* 2012, **24**:842–857.

44. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

45. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL: **Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae.** *Plant Physiol.* 2010, **153**:1398–1412.

46. Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S, Rodríguez-Concepción M, Martínez-García JF: **Interaction of shade**

avoidance and auxin responses: a role for two novel atypical bHLH proteins. *EMBO J.* 2007, **26**:4756–4767.

47. Cifuentes-Esquivel N, Bou-Torrent J, Galstyan A, Gallemí M, Sessa G, Salla Martret M, Roig-Villanova I, Ruberti I, Martínez-García JF: **The bHLH proteins BEE and BIM positively modulate the shade avoidance syndrome in Arabidopsis seedlings.** *Plant J.* 2013, **75**:989–1002.

48. Wang H, Zhu Y, Fujioka S, Asami T, Li J, Li J: **Regulation of Arabidopsis brassinosteroid signaling by atypical basic helix-loop-helix proteins.** *Plant Cell* 2009, **21**:3781–3791.

49. Tsukagoshi H, Busch W, Benfey PN: **Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root.** *Cell* 2010, **143**:606–616.

50. He YJ1, McCall CM, Hu J, Zeng Y, Xiong Y. **DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases.** *Genes Dev.* 2006, **21**:2949-2954.

51. Zhang Y, Feng S, Chen F, Chen H, Wang J, McCall C, Xiong Y, Deng XW: **Arabidopsis DDB1-CUL4 ASSOCIATED FACTOR1 forms a nuclear E3 ubiquitin ligase with DDB1 and CUL4 that is involved in multiple plant developmental processes.** *Plant Cell* 2008, **20**:1437–1455.

52. Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M: **Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice.** *Plant Physiol.* 2006, **141**:924–931.

53. Mori M, Nomura T, Ooka H, Ishizaka M, Yokota T, Sugimoto K, Okabe K, Kajiwara H, Satoh K, Yamamoto K, et al.: **Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis.** *Plant Physiol.* 2002, **130**:1152–1161.

54. Nakagawa H, Tanaka A, Tanabata T, Ohtake M, Fujioka S, Nakamura H, Ichikawa H, Mori M: **Short grain1 decreases organ elongation and brassinosteroid response in rice.** *Plant Physiol.* 2012, **158**:1208–1219.

55. Yu H, Braun P, Yildirim M a, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al.: **High-quality binary protein interaction map of the yeast interactome network.** *Science* 2008, **322**:104–110.

56. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol. Biol. Evol.* 2013, **30**:772–780.

57. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood**

phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 2010, **59**:307–321.

58. Huson DH, Scornavacca C: **Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks.** *Syst. Biol.* 2012, **61**:1061–1067.

6.3 SnRK1, a potential yield enhancing gene linked to source/sink distribution

Importance of source/sink distribution for yield

Plants are autotrophic organisms for carbon assimilation, which implies that they assimilate or 'eat' carbon dioxide during photosynthesis in chloroplasts and subsequently allocate the assimilated carbon throughout the plant. Two types of organs can be distinguished: net photoassimilate exporters, mainly mature green 'source' leaves and net importers, the 'sinks' e.g., roots, flowers, and storage organs like seeds. The general route of photoassimilates can be briefly described as follows. Upon fixation of carbon dioxide in the chloroplasts of mesophyll cells, triose phosphates either enter the cytosol for mainly sucrose formation or remain in the stroma to form transiently stored starch which is degraded during the night and enters the cytosol as maltose or glucose to be further metabolized to sucrose. In both cases, sucrose enters the phloem for long distance transport or is transiently stored in the vacuole. In the majority of plant species, sucrose is actively loaded into the phloem through the apoplast. Following long distance transport, it is released into sink organs through plasmodesmata (symplastic transport) or the cell wall (apoplastic transport), where it enters cells providing a source of carbon and energy. In storage organs, sucrose is stored as such or transformed to starch for storage in plastids, to oil in oil bodies, or – in combination with nitrogen – to protein in protein storage vacuoles and protein bodies.

Apart from autotrophic, plants are also sessile organisms. Integration of environmental signals with local sink establishment, carbon metabolism and sugar accumulation is thus particularly important. For example, to build a healthy body, plants need 25-30 chemical elements other than carbon, oxygen and hydrogen. As the availability of these elements (except for nitrogen in cultured crops) is finite per unit land area, plants have to compete for sufficient uptake. In addition, different types of stress can result in energy deprivation. Indeed, a reduction in photosynthesis and/or respiration is often associated with stress, in turn resulting in energy deprivation and ultimately in growth arrest or even cell death [1,2]. Stress could thus be partly decoded as an energy-deficiency signal that triggers convergent responses independently of the origin of its cause [3]. Hence, plants need to co-ordinate carbon assimilation to what their environment permits. Co-ordinating energy and metabolic homeostasis is thus a major challenge, and an intimate relationship exists between energy availability and nutrient availability, stress tolerance, survival, cell growth and longevity [1]. Summarized, although traditionally associated with sugar deprivation and darkness, energy deficit is, to varying degrees, probably also triggered by all adverse conditions that impinge on cellular energy and metabolite levels. Based on recent findings, a role of energy signalling mediated by SUCROSE NON-FERMENTING RELATED KINASE 1 (SnRK1) in the orchestration of transcriptional and post-translational responses was suggested [1,4,5]. From these findings, the view emerged that SnRK1 is a major regulatory switch that integrates various nutrient and metabolic signalling pathways, regulating energy and stress responses [6].

The SnRK1-family of protein kinases.

The Ser/Thr protein kinase family of SUCROSE NON-FERMENTING RELATED KINASE 1 proteins is structurally and functionally analogous to its yeast and mammalian counterparts, SUCROSE NON-FERMENTING (SNF1) and AMP-ACTIVATED PROTEIN KINASE (AMPK) respectively [7]. Similar to their yeast and mammalian counterparts, SnRK1 proteins function as fuel gauge sensors that sense cellular carbohydrate status and/or AMP/ATP levels in order to maintain growth in response to available energy [6,8–10]. As such, SnRK1 kinases regulate global metabolism and energy status of the plant, for example in response to low glucose/high sucrose levels, dark period, hypoxia, salinity and/or pathogen or herbivore attack [5]. They control metabolic and signalling pathways at the post-translational and transcriptional levels respectively, leading to modulation of nitrogen, sucrose, and lipid metabolism, organogenesis, and senescence [4,8].

SnRK1 proteins participate as the catalytic (α) subunit in heterotrimeric protein complexes, further composed of two regulatory subunits β- and γ [10] (Figure 1a). In plants, γ-related subunits homologous to the yeast activating SNF4 subunit can be grouped in three classes, KINβγ [11,12], KINγ [13] and the PV42/BsnIP1-type [14] proteins. They contain four in-tandem cystathionine β-synthase (CBS) motifs that function as dimers to form two domains constituting the site of fixation of the regulatory AMP or ATP molecules in mammals [15,16] (Figure 1a). The plant-specific AKINβγ-types of subunits resulted from the fusion between a γ-type protein and a carbohydrate-binding module (CBM) of β-subunits [12]. This CBM can mediate the interaction of AKINβγ with two proteins that are related to plant-pathogen interactions [17]. Curiously, nevertheless this extension AKINβγ proteins complement snf4 yeast mutants [11] unlike classical AKINγ-subunits [13] and are therefore considered the 'true' orthologs of SNF4 [18].

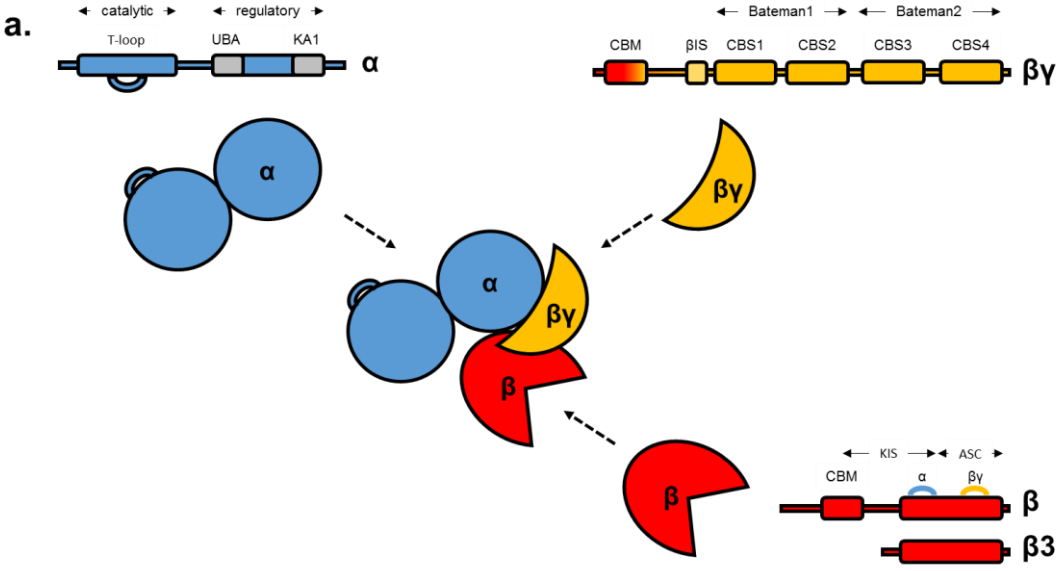
The β-subunits include three domains and mediate the interactions between the α- and γ-subunits within the heterotrimeric complex (Figure 1a). The functions of these three domains were elaborated in studies from *Saccharomyces cerevisiae* and mammals. The association-to-the-complex (ASC) domain located at the C-terminus allows interaction of the β-subunit with the γ- and α- subunits, enabling its role as scaffold [19,20]. An internal kinase-interacting sequence (KIS) designates the region comprising the interaction site of the β-subunits with the kinase subunit and the carbohydrate-binding module (CBM) [19]. The CBM shows characteristics of an N-isoamylase domain and binds in some cases to glycogen in mammals [21]. Although this module was found in the β-subunits of all three kingdoms, its function in plants remains unclear [22]. Plant β-subunits can be subdivided into two classes [13]. The first class including Arabidopsis KINβ1 and KINβ2 shares all the above mentioned characteristics with yeast and mammalian β-subunits. The other class of atypical KINβ3-subunits is truncated at the N-terminus and therefore lacks the CBM domain [23]. Despite this deletion, one of these atypical proteins complemented the yeast triple β-subunit mutant, suggesting that some basic functions have been conserved [23].

The SnRK1 α-subunit contains an N-terminal protein kinase catalytic domain and a C-terminal regulatory domain (Figure 1a). The kinase domain contains an

activation loop (T-loop), but phosphorylation of this T-loop did not seem to be essential for SnRK1 activity [4]. The regulatory domain interacts with the γ- and/or βγ-subunits [24] and contains a ubiquitin-associated (UBA) and a kinase-associated 1 (KA1) domain. The former could mediate interaction with ubiquitinated proteins [25], whereas the latter is responsible for interaction with regulatory subunits and upstream phosphatases [11]. The Arabidopsis SnRK1 family consists of three members, *SnRK1.1/AKIN10*, *SnRK1.2/AKIN11*, and *SnRK1.3/AKIN12*, of which the latter is probably a pseudogene [1]. The AKIN10 and AKIN11 kinases were identified as central regulators of the transcriptome in response to darkness and multiple types of stress signals [1]. Overexpression of *AKIN10* in Arabidopsis leads to metabolic re-programming with resulting delay in flowering and senescence of mature plants [1]. Whereas *AKIN10* is broadly expressed, *AKIN11* expression is spatially restricted [26]. Also in contrast to *AKIN10*, overexpression of *AKIN11* resulted in early rather than delayed flowering [26].

Rice contains three SnRK1 genes, which are classified into the SnRK1a (*OSK1*) and SnRK1b (*OSK24* and *OSK35*) sub-families, based on amino acid sequence similarities as well as expression patterns [27,28] (Figure 1b). The SnRK1a sub-family is more closely related to the homologs present in dicots, whereas the SnRK1b subfamily is unique to cereals [29]. The conserved function of SnRK1 protein kinases in rice was demonstrated in the sugar starvation signalling cascade in growing seedlings [30]. Studies on rice embryos indicated that SnRK1a/OSK1 acts upstream of the MYBS1 transcription factor to induce the α-amylase gene *αAMY3* during the early stages of germination to nourish the embryo through degradation of the starchy endosperm [30]. Therefore it plays a key role in regulating seed germination and seedling growth in rice. This process is of particular importance under conditions of extreme sugar starvation like hypoxia [30]. It indeed seems that OSK1 activity is determinant for the tolerance of some varieties of young rice seedlings to flooding [31]. *OSK1* is thought to play a broader role in sugar role than SnRK1b's, as *OSK1* is uniformly expressed in various growing tissues, including young roots and shoots, flowers and immature seeds [27]. In contrast, the *OSK24* and *OSK35* are preferentially expressed in the caryopsis (the rice grain) [28]. It is therefore tempting to speculate that SnRK1bs play some role in the starch accumulation in cereals. Especially the spatial and developmental patterns of the *OSK24* gene expression appear to be closely related to sucrose and starch metabolisms in the developing caryopsis and the leaf sheath where sink to source transition occurs [28].

Figure 1. Components of plant SnRK1 heterotrimeric complexes.



b.

	α-subunit			β-subunit			βγ-subunit	
Yeast		SNF1			SIP1/SIP2/GAL83		SNF4	
Human		AMPK			AMPKβ		AMPKγ	
Arabidopsis		AKIN10	at3g01090		AKINβ1	at5g21170	AKINβγ	at1g09020
		AKIN11	at3g29160		AKINβ2	at4g16360	(AKINγ	at3g48530)
				truncated β	AKINβ3	at2g28060		
Rice	SnRK1a	OSK1	loc_os05g45420		KINβ1	loc_os05g41220	KINβγ1	loc_os03g63940
	SnRK1b	OSK24	loc_os08g37800		KINβ2	loc_os03g20340	KINβγ2	loc_os04g32880
		OSK35	loc_os03g17980		KINβ4	loc_os07g48790		
				truncated β	KINβ3	loc_os09g20010		

a. Composition of the heterotrimeric SnRK1 complex and structure of the different subunits. The α-subunit (blue) contains a catalytic domain and a regulatory domain. The catalytic domain encompasses a regulatory T-loop, whereas the regulatory domain possesses a ubiquitin-associated (UBA) and a kinase-associated (KA1) domain for binding the β- and βγ-subunits. The βγ-subunit (yellow) has two 'Bateman', each built from two cystathionine-β-synthase (CBS) domains, and a β-interacting sequence (βIS). Resulting from domain fusion throughout evolution, the βγ-subunits acquired a carbohydrate-binding module (CBM, orange) at the N-terminus. Nonetheless, they were the only plant γ-subunits found to complement the snf4 mutant phenotype in yeast and are therefore considered 'true' orthologs of the yeast and mammalian γ-subunits. The β-subunit (red) harbours an association-to-the-complex (ASC) domain, containing the sites of interaction with γ- and α-subunits, a CBM and an N-terminal extension. The kinase interacting sequence (KIS) encompasses the region comprising CBM and interaction site with the α-subunit. KINβ3-subunits are atypical, as they lack the N-terminus including the CBM domain. **b.** SnRK1 subunits encoded in the Arabidopsis and rice genomes, and their relation to yeast and mammalian SNF1 and AMPK subunits respectively. Rice SnRK1α-subunits can be subdivided in a SnRK1a and cereal-specific SnRK1b subfamily, based on sequence similarities and expression patterns.

Regulation of and by SnRK1 complexes

SnRK1 is a master regulator of metabolism and transcription in response to energy deprivation and abscissic acid (ABA) signals, and is inactivated by sugars that restore energy balance [4,32,33]. At the post-translational level, SnRK1 directly regulates substrates varying from key metabolic enzymes such as sucrose phosphate synthase, nitrate reductase and 3-hydroxy-3-methylglutaryl-coenzyme

A reductase (HMG-CoA reductase) [7], over transcription factors like FUSCA3 [34], to factors regulating cell cycle progression [35] (Figure 2b). Indirectly, SnRK1 activates adenosine diphosphoglucose pyrophosphorylase (AGPase, the enzyme catalysing the first committed step in starch synthesis) in response to high sucrose levels, potentially through modulation of its redox-status [36]. In addition, SnRK1 is able to induce major shifts in gene expression [1] (Figure 2b). More specifically, Arabidopsis KIN10 promotes catabolism by regulating expression of a broad array of genes involved in a variety of major catabolic pathways that provide alternative sources of energy and metabolites. Conversely, a large set of genes involved in the energy-consuming ribosome biogenesis and anabolism are co-ordinately repressed by KIN10 [1]. SnRK1s control a large number of genes encoding transcription factors, chromatin remodelling proteins, and a plethora of signal transduction components [4]. Besides their role in metabolism and stress responses these kinases regulate virtually all aspects of cell function as well as multiple developmental processes. For example, SnRK1 showed crucial in seed filling and maturation, and to affect embryo development and cotyledon growth in pea [37,38], as well as pollen development in barley [39], and lateral organ development and phase transition in Arabidopsis [34] (Figure 2b). Strict regulation of the SnRK1 complex's activity is therefore crucial to maintain normal plant functioning and development. Different layers of SnRK1 regulation exist and are illustrated in Figure 2a.

A first layer of regulation is provided by the subunit availability and composition of the SnRK1 heterotrimers. Since each subunit of the SnRK1 complex is represented by different members in plants, several assemblies are possible *in vivo.* These add up to 12 and 24 options in Arabidopsis and rice respectively. Also, oligomerisation between individual heterotrimers is possible and can influence the kinase activity. Oligomerisation of SNF1 heterotrimers were observed in the yeast and mammalian field [40–43]. How this oligomerisation impacts kinase activity is not yet fully investigated. For mammalian AMPK, formation of higher order oligomers was associated to an inactive state of the complex which upon activation would disassemble into dimeric and monomeric units of the heterotrimeric complex [43]. On the other hand, the activation loop of SNF1 becomes inaccessible for phosphorylation by upstream kinases when the catalytic subunits form dimers, suggesting that also the dimeric forms of heterotrimers would be inactive [44].

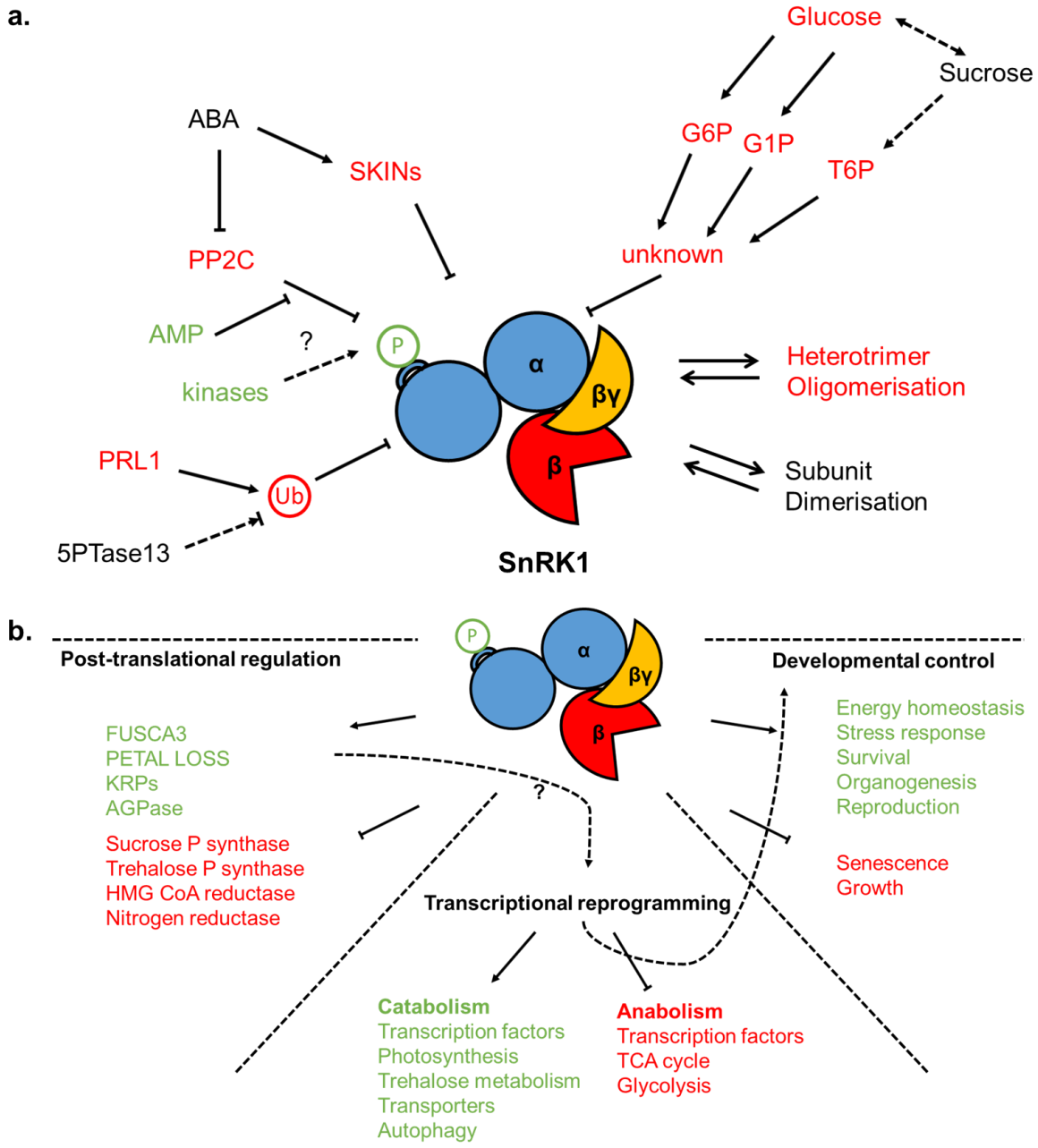
Yeast and mammalian α-subunits require phosphorylation of a threonine residue in the T-loop for kinase activity. In plants, this could not be unambiguously determined [1]. The phosphorylation status of the T-loop can be altered by antagonising kinases and phosphatases. SnRK1 activating kinases (SnAKs) in Arabidopsis [45] and CIPK15 (CALCINEURIN B-LIKE-INTERACTING PROTEIN KINASE 15) in rice [46] were proposed as potential upstream regulators, but for none of these *in vivo* biochemical validation could be provided. As antagonists of T-loop phosphorylation, two protein phosphatase 2C phosphatases (PP2C), named ABI1 (for ABA INSENSTIVE 1) and PP2C1 were proposed to dephosphorylate SnRK1 [33]. These phosphatases are well known negative regulators of the ABA pathway. Upon ABA perception, ABA receptors block the repressive actions of ABI1 and PP2C1, thereby activating SnRK1 activity [47].

Remarkably, ABA also represses SnRK1 signalling, through SNRK1-INTERACTING NEGATIVE REGULATORS (SKINs) during germination and early seedling growth [48]. This may relate to the suggestion that the effect of ABA may differ between source and sink tissues in a similar manner as animal hormones control AMPK in opposite manner in different tissues [49].

Apart from kinase activity, also protein stability is a potential layer of regulation. In agreement with this, AKIN10 is targeted for proteasomal degradation under low nutrient conditions through interplay between two WD40-repeat domain containing proteins [50]. Both myo-inositol polyphosphate 5-phosphatase (5PTase13) and PRL1 (for PLEITROPIC REGULATOR LOCUS 1) interact with AKIN10 and regulate the delivery of AKIN10 to the CUL4-DDB1 complex for proteasomal degradation [51,52]. It was suggested that 5PTase13 and PRL1 have opposing functions regarding AKIN10 stability under low-nutrient and sugar-stress conditions. In this model, 5PTase13 acts as a positive regulator of AKIN10 under low-nutrient or sugar-stress conditions by reducing the amount of the kinase targeted for proteasomal degradation. In contrast, PRL1 acts as a negative regulator and facilitates its destruction [50].

Although T-loop phosphorylation and ubiquitination are the best studied mechanisms of regulating activity, several other post-translational modifications are known from yeast and mammals, including acetylation, SUMOylation and myristoylation [49]. These still have to be fully established for plant SnRK1s however. Apart from regulation at the post-translational or complex assembly level, SnRK1 activity is further influenced by AMP/ATP ratio and sugar phosphates such as glucose-1-phosphate, glucose-6-phosphate and trehalose-6-phosphate. Whereas the adenylates probably function through AMP by protecting the T-loop from dephosphorylation [53], the sugar phosphates are acting via an intermediary factor that is separable from SnRK1 [54].

Figure 2. The SnRK1 complex regulates and is regulated.



a. Overview of the regulatory mechanisms controlling SnRK1 activity. Factors negatively regulating SnRK1 function are marked red, whereas positive regulators are depicted green. ABA and 5PTase have a more ambiguous control over SnRK1, depending on the stress and energy status of the cell and the developmental context. ABA: abscissic acid; SKINs: SnRK1A-interacting negative regulators; PP2C: phosphatase 2C; AMP: adenosine monophosphate; PRL1: pleiotropic regulatory locus 1; 5PTase13: myo-inositol polyphosphate 5-phosphatase; G6P: Glucose-6-phosphate; G1P: Glucose-1-phosphate; T6P: Trehalose-6-phosphate; P: phosphorylation; Ub: ubiquitination. **b.** The SnRK1 master regulator controls multiple biological levels. Activated components are marked green, repressed components red. FUS3: B3-domain transcription factor FUSCA3; KRP: KIP-related protein; P: phosphate; HMG CoA reductase: 3-hydroxy-3-methylglutaryl-coenzyme A reductase; TCA: tricarboxylic acid cycle

Established protein-protein interactions for SnRK1

Yeast two hybrid experiments using the rice SnRK1 proteins OSK1 and OSK24 resulted in the identification of respectively 24 and 26 potential interaction partners [55]. A representation of the OSK1-interactions is shown in Figure 4f. The expected interactions between these α-subunits and the β-subunits loc_os05g41220 and loc_os09g20010 were confirmed. The latter is a member of the β3-subfamily that lacks the N-terminal region as mentioned before. Both bait SnRKs also interacted with loc_os03g63940, belonging to the class of βγ-subunits and loc_os02g38780, a putative 2C protein phosphatase (PP2C). This phosphatase could be involved in regulation of the phosphorylation status of the T-loop of the two SnRK1's [55] (Figure 2a).

Consistent with the observations from the Y2H experiments, the βγ-subunit loc_os03g63940 was identified in a complex with OSK1 in TAP tag analyses on rice leaves [56]. Interaction data from this study for OSK1 are shown in Figure 4e. OSK1 and OSK35, another α-subunit, were used as bait protein. In fact, both baits were found associated with the two βγ-subunits in rice, loc_os03g63940 and loc_os04g32880 [56]. Also interactions between two β-subunits loc_os0920010 and loc_os05g41220, and OSK1 were confirmed, whereas OSK35 rather associated with β-subunit loc_os05g41220. In addition, both SnRK1α subunits co-purified the β-subunit loc_os03g20340. Curiously, the TAP experiments did not detect any γ-subunits. Instead, OSK24 was identified in association with OSK35. This finding, together with the detecting of both βγ- and β-subunits in the same protein complex, led to the author's suggestion that different SnRK1 heterotrimers could associate *in vivo* [56].

Tandem affinity purification of Arabidopsis AKIN10

Similar to the experimental set-up for Ta_HLH, the aim of purifying interactors from AKIN10 was twofold. First we wanted to obtain a molecular view on the action mechanism of the AKIN10 gene in rice. Second, we were seeking for further validation by comparing the output from our workflow with the interaction dataset resulting from an earlier screen with AKIN10 using Arabidopsis cell suspension cells by the Functional Interactomics group at the 'Vlaams Instituut voor Biotechnologie' according to the protocol of Van Leene et al., 2008 [57]. As the SnRK1 complex thus was purified from both Arabidopsis and rice [56], this was a feasible opportunity to verify whether a homologous protein is able to participate in endogenous complexes. We selected AKIN10 rather than AKIN11 as the latter had a more restricted expression pattern [26], which might be less suitable in our whole seedling-based screening protocol. Also, it was shown that SnRK1 activity in Arabidopsis cells is mostly a function of the AKIN 10 gene product, thus the AKIN 11 gene may play a rather restricted role in regulating most plant metabolism, stress and/or energy sensing [58].

We TAP-tagged and affinity-purified the Arabidopsis α-subunit of the SnRK1 complex AKIN10. As a sensor and signal transducer for deprivation of sugar and energy, AKIN10 targets a broad array of genes that orchestrate transcription networks, promoting catabolism and suppressing anabolism [1]. We designed an

N-terminal fusion of AKIN10 with the GS tag expressed by PRO_{35S} and a C-terminal fusion with the GSgreen tag expressed under control of PRO_{GOS2}. These translational fusions were introduced in rice callus, T0 shoots and T1 seedlings. Expression of the TAP constructs in the different transgenic tissues was verified at the protein level before engaging to the actual purification experiments (Figure S2).

We executed in total eleven purifications on the three types of material. From transgenic callus cells expressing the C-terminal GS-tagged bait, we performed two replicate experiments on 50mg total protein. Transgenic callus expressing the GSgreen-tagged bait was used for two replicate experiments on 25 mg and on 50mg total protein. In parallel, we screened the bait constructs in plant tissues. T0 shoots, immediately generated from the transgenic callus cells described above, were employed for nonreplicate experiments on the N-terminal fused bait. Transgenic seeds from 60 independent transgenic events were sown to grow 2 weeks old seedlings expressing the baits for duplicate purifications on both TAP fusions. We could further compare our results with interaction data from previous experiments using C- and N-terminal fusions of AKIN10 to the GS TAP tag expressed under the control of the 35S promoter in Arabidopsis cell cultures (Figure 3).

Figure 3. Interaction data from TAP experiments on the Arabidopsis SnRK1 α-subunit AKIN10 in rice.

				T0 callus			T0 shoots	T1 seedlings						
	accession	description	symbol	PRO_{35S}:GS-AKIN10 (2)	PRO_{GOS2}:AKIN10-GSgreen (2)	PRO_{GOS2}:AKIN10-GSgreen* (2)	PRO_{35S}:GS-AKIN10 (1)	PRO_{35S}:GS-AKIN10 (2)	PRO_{GOS2}:AKIN10-GSgreen (2)	total	ara	Y2H	rice	target
SnRK1 core complex constituents	loc_os05g45420	SnRK complex α-subunit, SnRK1a	OSK1	2	2	2	1	2	2	11	x	x	x	
	loc_os08g37800	SnRK complex α-subunit, SnRK1b	OSK24		2					2				
	loc_os03g17980	SnRK complex α-subunit, SnRK1b	OSK35				1			1				1
	loc_os03g20340	SnRK complex β-subunit 2	KINβ2	2	2	2	1	2	2	11	x			1
	loc_os05g41220	SnRK complex β-subunit 1	KINβ1	2	2	2		2	2	10		x		
	loc_os09g20010	SnRK complex β3-subunit	KINβ3	1		2		2	2	7	x	x	x	1
	loc_os03g63940	SnRK complex βγ-subunit 1	KINβγ1	2	2	2	1	2	2	11	x	x	x	1
	loc_os04g32880	SnRK complex βγ-subunit 2	KINβγ2	2	2	2	1	2	2	11			x	2
potential SnRK1 repressors	loc_os08g40510	SNRK1A-INTERACTING NEGATIVE REGULATOR 2	SKIN2			2				2	x			
	loc_os05g44900	OXIDATIVE STRESS 3-like 4	OsO3L4			2				2		x		1
	loc_os02g13370	OXIDATIVE STRESS 3-like 3	OsO3L3	1						1				2
	loc_os02g11830	potential PLEIOTROPIC REGULATORY LOCUS 1	OsPRL1-like						1	1				
potential SnRK1 targets	loc_os03g28330	sucrose synthase					1	2	2	5				4
	loc_os06g09450	sucrose synthase						1		1				4
	loc_os01g53000	trehalose synthase		2			1	2		5				
	loc_os01g54560	trehalose synthase		2			1	2		5				1
	loc_os02g54820	trehalose synthase					1	2		3				2
	loc_os05g44100	trehalose synthase						2		2				1
	loc_os09g23350	trehalose synthase						2		2	x			1
	loc_os03g12360	trehalose synthase					1			1				4
	loc_os08g25390	bifunctional aspartokinase/homoserine dehydrogenase						1		1				4
	loc_os02g51710	lypoxygenase		2						2				
	loc_os06g14240	hsp20/alpha crystallin family protein						2		2				
	loc_os02g44990	MIRNA's REGULATED AND ABIOTIC STRESS INDUCED F-BOX	MAIF1	1						1				

Interaction table for proteins retrieved from TAP experiments from rice T0 callus, T0 shoots and T1 seedlings. For GSgreen constructs, two replicate purifications were performed on 25mg and 50mg (asterisk) of total protein from transgenic callus. The number of replicate experiments for a construct are shown between brackets. The rice α-subunits are only accounted for if – in addition to our normal standards- at least one peptide uniquely matched the sequence of that specific rice protein and not AKIN10. Protein identifications that were not reproducibly found are shown in grey. Potential interactors that were retrieved in multiple bait groups and normally would have been assigned as background are marked red. ara: marked when the interactor was found in TAP experiments from N- or C-terminal GS-tagged AKIN10 in Arabidopsis cell suspensions. Y2H: marked when the interaction was confirmed in yeast-two hybrid interaction screening of rice kinases [55]. rice: marked when the protein was found to co-purify with TAPi-tagged OSK1 in rice leaves [56]. Target: illustrates how many times a SnRK1 substrate recognition site was found in the protein's sequence.

From the eleven purifications that were performed, we recurrently found multiple peptides matching with not only the AKIN10 protein, but also with OSK1, OSK24 and/or OSK35. Since the AKIN10 sequence was not present in the database we used for searching, we could not unambiguously show that the bait was identified.

In all purifications, we could detect peptides uniquely matching to OSK1 and not AKIN10, whereas OSK24 and OSK35 were unambiguously observed in callus and shoot purifications respectively (Figure 3). This hints towards co-purification of these protein with the AKIN10 bait. Rice SnRK1 α-subunits from the b-type were previously shown to co-precipitate [56]. This finding made the authors suggest that SnRK1 heterotrimers are associating *in vivo*. Our results are in line with that finding and hint towards association of different SnRK1 heterotrimers containing SnRK1 α-subunits from both the a- and b-type. As mentioned, oligomerisation of SNF1 heterotrimers and/or dimerization of α-subunits were previously observed in the yeast and mammalian field. It is however not clear what the impact is on the kinase activity. Further supporting the general believe that SnRK1s function as heterotrimers, we found all three β- and two βγ-subunits present in the rice genome co-purified. This thus implies that the Arabidopsis AKIN10 is able to at least participate in endogenous rice complexes. Similar as what was observed previously, no γ-subunits were detected. Accordingly, it was indicated from a phylogenetic study that in fact plant KINβγ-subunits are the true orthologs of γ-subunit genes from fungi and animals [18].

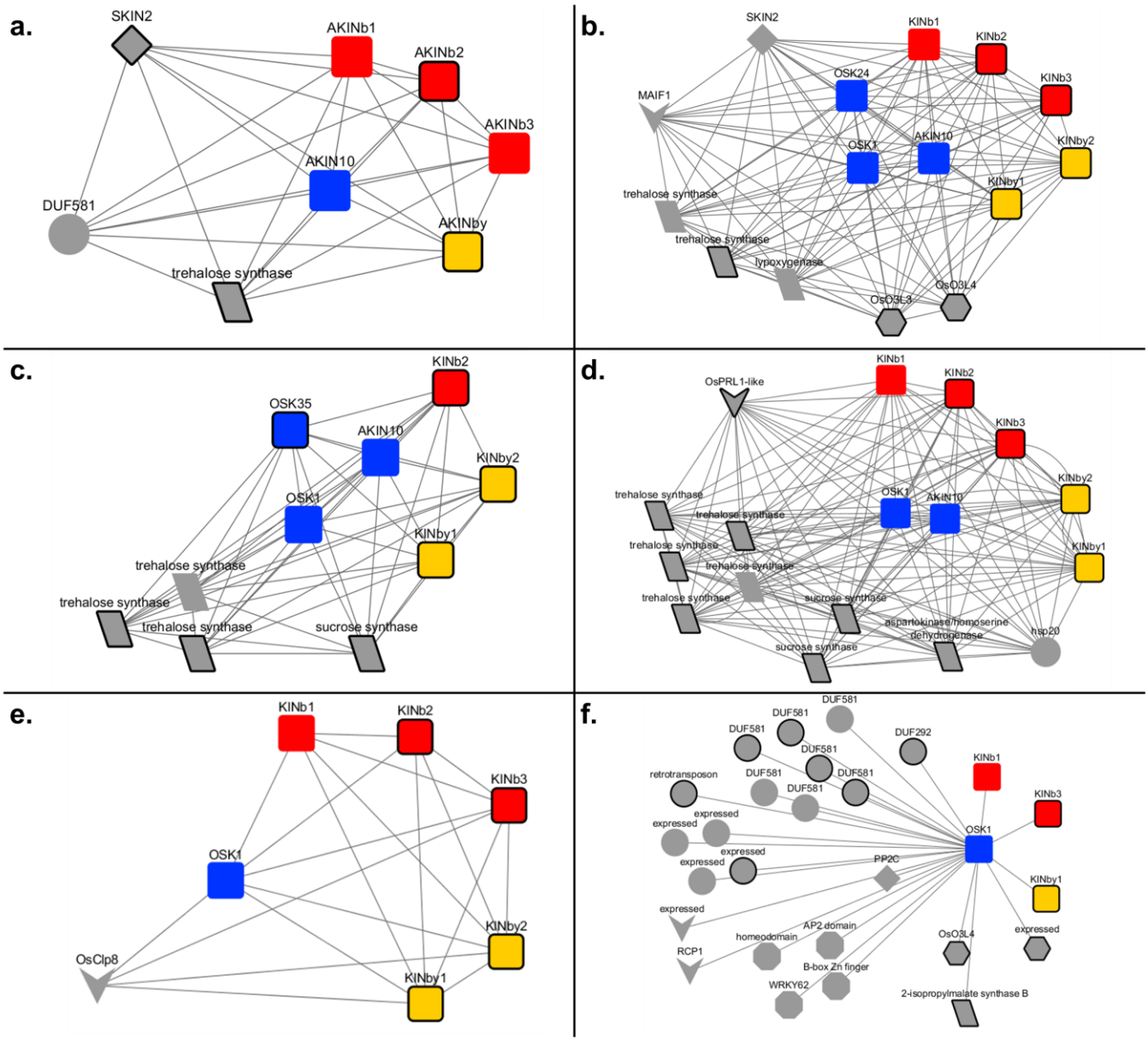
Apart from the core constituents of the SnRK1 complex, we retrieved two known repressors of SnRK1 activity. We detected a SnRK1-interacting negative regulator 2 (SKIN2) [48] and its potential Arabidopsis ortholog in rice callus and Arabidopsis cell suspension cultures respectively. Interaction of SKIN2 with OSK1 was initially detected through Y2H and confirmed in rice embryos [48]. SKIN proteins contain a distinct KSD domain that is required for antagonizing OSK1 function [48]. During germination and early seedling growth, ABA shifts SKIN2 localisation from nucleus to cytoplasm, where it binds and represses OSK1 function [48]. It is thought that in this pathway OSK1 regulates source-sink communication during seedling growth, whereas under abiotic stress ABA – through SKINs - antagonizes the function of OSK1 to restrict seedling vigor [48]. We found a second SnRK1 repressor in a rice protein homologous to PRL1. Arabidopsis PRL1 potentially mediates AKIN10 degradation by acting as the substrate receptor of a CUL4-based E3 ubiquitin ligase [52].

To verify whether some of these co-purifying proteins could be genuine targets of AKIN10, we checked for presence of the kinase recognition motif defined [7]. The phosphorylation of sucrose synthase by SnRK1 was already established for plants. We found various members of the trehalose phosphate synthase (TPS) family as potential substrates. A previous study already suggested *in vitro* phosphorylation of Arabidopsis thaliana TPS isozymes containing the SnRK1 consensus recognition [59]. Trehalose-6-phosphate (T6P) is known to regulate (inhibit) SnRK1 activity through an intermediary factor. It is therefore plausible that SnRK1 controls T6P metabolism through regulation of its metabolic enzymes providing a feedback mechanism. Another metabolism related target was found in a bifunctional aspartate kinase/homoserine dehydrogenase. These two activities catalyze the first and the third steps toward the synthesis of the essential amino acids threonine, isoleucine and methionine and hint that SnRK1 is controlling these committed steps in synthesis of these amino acids.

Also OsO3L4 and OsO3L3 (for *Oryza sativa* OXIDATIVE STRESS 3-LIKE 4 and 3 respectively) contained the SnRK1 phosphorylation motif. These are homologous to the Arabidopsis OXIDATIVE STRESS 3 (OXS 3) protein. OXIDATIVE STRESS 3 was discovered through Cadmium tolerance selection screening in *Schizosaccharomyces pombe*. Fission yeast overexpressing either the *Brassica juncea* homolog for OXS3 or the Arabidopsis OXS3 cDNA enhanced tolerance to a range of metals and oxidizing chemicals [60]. Analysis of Arabidopsis mutant and overexpression lines confirmed a role of OXS3 in oxidative stress tolerance [60]. The OXS3 family of proteins share a highly conserved domain corresponding to a putative N-acetyltransferase or thioltransferase catalytic site and OXS3 proteins localise in the nucleosome, in discrete parts of the chromatin. Therefore it was hypothesized that OXS3 might act as a chromatin remodelling factor to anticipate stress responses [60]. We found OsO3L4 and OsO3L3 in experiments from rice callus, for which OsO3L4 was confirmed through yeast two hybrid experiments [55]. The latter hints to direct interaction of the kinase subunit with the OXS-proteins. Remarkably, OsO3L4 and OsO3L3 (and OXS3-like proteins in general) contain the same distinct KSD domain that is necessary for SKIN proteins to antagonize OSK1 function (Figure S3). Therefore, OsO3L4 and OsO3L3 could be potential negative regulators of SnRK1.

Another interesting interactor found was MIRNA'S REGULATED AND ABIOTIC STRESS INDUCED F-BOX 1 (MAIF1). Although not found reproducibly, previous findings from rice hint toward involvement of MAIF1 in the SnRK1 pathway. MAIF1 expression was found induced by sucrose, ABA and abiotic stresses [61]. Further, *MAIF1* expression is induced during cell division of root tips and transgenic rice plants overexpressing the gene showed promoted root growth and reduced ABA sensitivity and abiotic stress tolerance [61]. This implies that under abiotic stress conditions, *MAIF1* overexpressors lose their root growth constraint.

Figure 4. Cytoscape representation of protein-protein interaction data retrieved from SnRK1 in rice and Arabidopsis.



Cytoscape networks based on interaction data retrieved from Arabidopsis cell suspension cultures (**a.**), rice T0 callus (**b.**), T0 shoots (**c.**), T1 seedlings (**d.**), 6-8 weeks old seedlings from a study on rice kinases [56] (**e.**), and yeast-two hybrid experiments on rice kinases [55] (**f.**). The SnRK1 components are illustrated as rounded rectangles in blue, yellow and red for α-, βγ-, and β-subunits respectively. Other symbols used are: hexagon: chromatin modifier; parallelogram: metabolic enzyme; diamond: potential SnRK1 regulator; V-shape: (proteasome-mediated) protein degradation; octagon: transcriptional regulator; oval: unknown. Proteins that have a SnRK1 potential substrate recognition site in their sequence have border in bold.

Conclusion

Plant growth and development are tightly controlled in response to environmental conditions that influence the availability of photosynthetic carbon in the form of sucrose. SnRK1 is an important regulation switch involved in controlling plant growth and development in response to carbon availability. Developmental processes that are regulated by SnRK1s range from embryo development to leaf senescence. SnRK1 is most likely involved in the adjustment of metabolism and growth in response to starvation. A potential yield benefit is demonstrated from transgenic tomato lines overexpressing apple rootstock SnRK1 [62]. These showed increased photosynthetic rate, higher starch content in leaves and higher red-ripening fruits compared to the wild type. The transgenic fruits also ripened ~10 days earlier compared to the wild type. Further, the transgenic lines had a more efficient N-uptake, suggesting that overexpressing SnRK1 can increase both the carbon and nitrogen assimilation rate of the plant as well as regulate development in fruit [62].

Energy status sensing is also of crucial importance in relation to seed yield. Grain weight is determined by grain size provided that the entire panicle is fully filled. Grain yield in rice can thus be defined as the product of filling efficiency [63]. This grain filling rate is characterized by the capability of carbohydrate accumulation in the leaves and stems (the source) and the translocation of assimilates from these source tissues to the grains (the sink) [64]. Members of the cereal-specific SnRK1b subfamily show distinct expression in the caryopsis and are considered to be associated with the development of sink tissue capacity in developing seeds [28,65]. They could therefore provide a potential target for improving seed yield.

We utilised our established TAP platform to identify potential interactors for the Arabidopsis α-subunit of the SnRK1 complex in rice. The strength of our workflow is illustrated by the retrieval of not only the core complex members, but also known regulators. In addition, we uncovered potential new regulators and even putative SnRK1 targets, as some interactors contained the kinase recognition motif for phosphorylation.

From retrieval of the expected β- and βγ-subunits, we could deduce that the Arabidopsis AKIN10 protein is able to participate in endogenous rice heterotrimeric complexes. Similar to observations from earlier experiments on the rice α-subunits of the SnRK1 complex [55,56], the bait seemed to interact with the other α-subunits. This was also observed for yeast SNF1 and mammalian AMPK subunits [43,44]. The physiological significance of this behaviour is however still unclear. Also in line what was earlier observed, we could not detect interaction with a γ-subunit. A potential cause suggested was the lack of an annotated γ-subunit in the rice protein databases [56]. Another possibility is that the βγ-subunits we did retrieve are the true orthologs for yeast *SNF4* [18], and that γ-subunits are simply not participating in *in vivo* SnRK1 complexes.

Apart from the core complex, we were able to retrieve two known negative regulators of SnRK1 activity. SKIN2 is regulated by ABA and retains SnRK1 in the cytoplasm upon stress, thereby preventing SnRK1 to enter the nucleus and trigger responses for mobilisation of nutrients [48]. The other repressor is a potential rice

homolog of PRL1 and functions in targeting SnRK1 for proteolysis through the E3 ubiquitination pathway [52]. Further, we found OXS3-like proteins as novel potential SnRK1 regulators since these also contain a KSD domain necessary for SKIN proteins to antagonise SnRK1 activity [48]. It is therefore plausible that OXS3-like proteins, similar to SKINs, negatively regulate SnRK1. OXS3-like proteins may on their turn be regulated by SnRK1, as both interactors contained potential SnRK1 phosphorylation sites. Previous studies presumed the OXS3 proteins to be chromatin remodelling factors that anticipate stress responses [60]. Also SNF1 and AMPK are both involved in regulating chromatin remodelling. AMPK can phosphorylate histone H2B, thereby activating transcription in response to stress [66], while SNF1 can activate gene expression by histone H3 acetylation, probably through recruiting the acetyl-transferase GCN5 (for GENERAL CONTROL OF AMINO ACID SYNTHESIS 5) [67]. Finding two proteins that could be linked to chromatin remodelling in our interaction dataset is in line with these observations.

Through presence of the kinase substrate motif for SnRK1 in their amino acid sequence, we further tagged two sucrose phosphates, six trehalose-phosphate synthases and a bifunctional aspartate kinase/homoserine dehydrogenase as potential SnRK1 targets. While sucrose synthases are known SnRK1 substrates, and phosphorylation Arabidopsis TPS isoforms was already confirmed *in vitro*, the bifunctional aspartate kinase/homoserine dehydrogenase and the earlier mentioned OXS3-like proteins are new and yet uncharted SnRK1 substrates. This hints that our protocol is sensitive enough to retrieve potential substrates of the SnRK1 kinase.

Unfortunately, although SnRK1 is known to regulate some transcription factors as FUS3 [68] and PETAL LOSS [69] in Arabidopsis and possibly MYBS1 in rice [30], we could not detect any transcriptional regulators in our TAP experiments. This in contrast to Y2H data from a kinase study, where multiple transcription factors were identified as potential interactors of OSK1 [55] (Figure 4f). One explanation could be that the interaction between SnRK1 and the transcription factors is too weak to survive the lengthy TAP protocol. Another possibility is that we simply didn’t assay the complex in the right physiological or developmental context. Indeed, the cDNA library used for Y2H analysis was based on a pool of RNA extracts from multiple conditions [55]. This greatly enlarges the chance to detect interactors available in any conditions, but also hampers interpretation of the results. In our system, both T0 callus, T0 shoots and T1 seedlings were grown on saturating levels of sucrose to sustain growth. It will be interesting to see if the complex constitution would change in relation to altered growth conditions.

Based on our filtering methods, some interactors would have been missed (Figure 2, proteins marked red). The only reason we could retrieve these proteins from the background list, is that their interaction with the SnRK1 complex was previously described. This illustrates the possible weakness of a subtraction list based on recurring proteins over independent purifications. Not only clear artefacts such as background proteins aspecifically binding tag or beads are included in the subtraction list, but also some *bona fide* interactors that simply bind to a broad range of proteins *in vivo*. Including quantitative measures such as normalised

spectral abundance factors could in this case help in indicating which frequently co-purifying proteins could be specific interactors in this case [70].

Materials & Methods

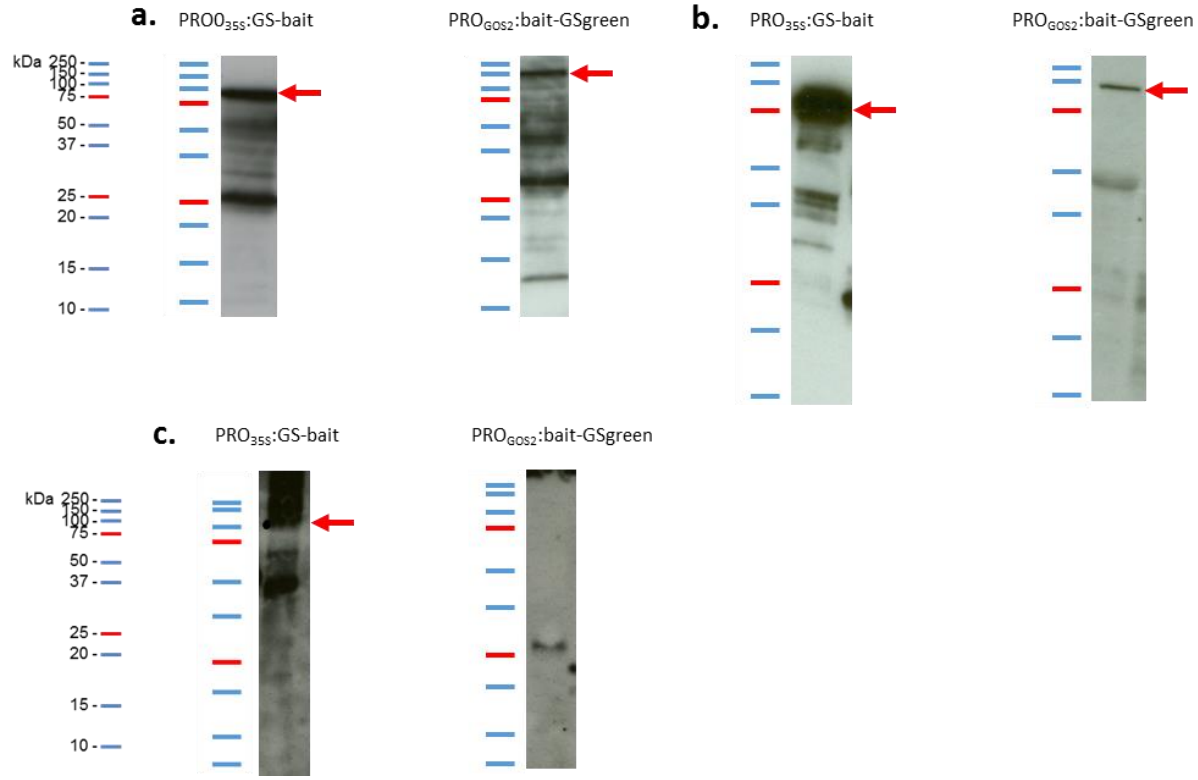
Screening of the AKIN10 bait in the rice TAP platforms
Screening of the bait in callus, T0 shoots and T1 seedlings was done as described in the materials and methods in chapter 6.1.

Supplementary information

Figure S1. Primers used in this chapter.

primer n°	name	primer sequence
prm27131	attB1 kozak SnRK1	ggggacaagtttgtacaaaaaagcaggctccaccatgttcaaacgagtagatgagttta
prm27147	attB2 SnRK1 no stop	ggggaccactttgtacaagaaagctgggtcgtttgaaggaggaaacttttgaagtg
prm27163	attB2r no start SnRK1	ggggaccactttgtacaagaaagctgggtggggacagctttcttgtacaaagtggccatgttcaaacgagtagatgagt
prm27179	attB3 SnRK1 stop	ggggacaactttgtataataaagttggtcagaggactcggagct

Figure S2. Expression analysis of the bait protein.



Protein extracts of the transgenic lines were analyzed by immunoblotting with antibodies against the GS tag or GSgreen tag to verify protein levels of the AKIN10 bait. **a.** Bait expression levels from callus extracts. **b.** Expression levels from T0 shoot extracts. **c.** Protein expression from T1 seedling extracts. Molecular marker values are provided in kilo Dalton (kDa) in the left panel. The molecular weight of the tagged proteins is 98.8 kDa for the GSgreen-tagged bait and 81.8 kDa for the GS-tagged bait.

Figure S3. SKIN2 and OXS3-like proteins share the KSD domain.

Sequence alignment of rice SKIN and OXS3-like proteins as described in [48] and [59]. The KSD domain, described by [48] is shown in highlighted with the red rectangle. Alignment was performed using the MAFFT tool [71], using the L-INS-I option. This iterative refinement search sustains global multiple sequence alignments and integrates strong local alignments to enrich for alignment of functional domains.

Supplementary background list and mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP).

The mass spec files for SnRK1 interaction data described in this chapter can be found in SI_Chapter_6_3_SnRK1_MS_data.xlsx.

Author contribution

The PhD candidate was in charge of generating the TAP constructs, the transformations, maintenance and upscaling of callus cell lines and plant material, analysis of the mass spectrometry data. He also wrote the manuscript. TAP purifications and mass spectrometry measurements were mainly done by employees of VIB.

References

1. Baena-González E, Rolland F, Thevelein JM, Sheen J: **A central integrator of transcription networks in plant stress and energy signalling.** *Nature* 2007, **448**:938–942.

2. Smith AM, Stitt M: **Coordination of carbon supply and plant growth.** *Plant. Cell Environ.* 2007, **30**:1126–1149.

3. Kenyon C: **The Plasticity of Aging: Insights from Long-Lived Mutants** *Cell* 2005, **120**:449–460.

4. Baena-González E, Sheen J: **Convergent energy and stress signaling.** *Trends Plant Sci.* 2008, **13**:474–482.

5. Halford NG, Hey SJ: **Snf1-related protein kinases (SnRKs) act within an intricate network that links metabolic and stress signalling in plants.** *Biochem. J.* 2009, **419**:247–259.

6. Hey SJ, Byrne E, Halford NG: **The interface between metabolic and stress signalling.** *Ann. Bot.* 2010, **105**:197–203.

7. Halford NG, Hey S, Jhurreea D, Laurie S, McKibbin RS, Paul M, Zhang Y: **Metabolic signalling and carbon partitioning: role of Snf1-related (SnRK1) protein kinase** *J. Exp. Bot.* 2003, **54**:467–475.

8. Halford NG, Paul MJ: **Carbon metabolite sensing and signalling.** *Plant Biotechnol. J.* 2003, **1**:381–398.

9. Rolland F, Baena-Gonzalez E, Sheen J: **Sugar sensing and signaling in plants: conserved and novel mechanisms.** *Annu. Rev. Plant Biol.* 2006, **57**:675–709.

10. Ghillebert R, Swinnen E, Wen J, Vandesteene L, Ramon M, Norga K, Rolland F, Winderickx J: **The AMPK/SNF1/SnRK1 fuel gauge and energy regulator: structure, function and regulation.** *FEBS J.* 2011, **278**:3978–3990.

11. Kleinow T, Bhalerao R, Breuer F, Umeda M, Salchert K, Koncz C: **Functional identification of an Arabidopsis snf4 ortholog by screening for heterologous multicopy suppressors of snf4 deficiency in yeast.** *Plant J.* 2000, **23**:115–122.

12. Lumbreras V, Alba MM, Kleinow T, Koncz C, Pagès M: **Domain fusion between SNF1-related kinase subunits during plant evolution.** *EMBO Rep.* 2001, **2**:55–60.

13. Bouly JP, Gissot L, Lessard P, Kreis M, Thomas M: **Arabidopsis thaliana proteins related to the yeast SIP and SNF4 interact with AKINalpha1, an SNF1-like protein kinase.** *Plant J.* 1999, **18**:541–550.

14. Slocombe SP, Laurie S, Bertini L, Beaudoin F, Dickinson JR, Halford NG: **Identification of SnIP1, a novel protein that interacts with SNF1-related protein kinase (SnRK1).** *Plant Mol. Biol.* 2002, **49**:31–44.

15. Scott JW, Hawley SA, Green KA, Anis M, Stewart G, Scullion GA, Norman DG, Hardie DG: **CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations** *J. Clin. Invest.* 2004, **113**:274–284.

16. Kemp BE: **Bateman domains and adenosine derivatives form a binding contract.** *J. Clin. Invest.* 2004, **113**:182–184.

17. Gissot L, Polge C, Jossier M, Girin T, Bouly J-P, Kreis M, Thomas M: **AKINbetagamma contributes to SnRK1 heterotrimeric complexes and interacts with two proteins implicated in plant pathogen resistance through its KIS/GBD sequence.** *Plant Physiol.* 2006, **142**:931–944.

18. Ramon M, Ruelens P, Li Y, Sheen J, Geuten K, Rolland F: **The hybrid four-CBS-domain KINβγ subunit functions as the canonical γ subunit of the plant energy sensor SnRK1.** *Plant J.* 2013, **75**:11–25.

19. Jiang R, Carlson M: **The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex.** *Mol. Cell. Biol.* 1997, **17**:2099–2106.

20. Vincent O, Carlson M: **Gal83 mediates the interaction of the Snf1 kinase complex with the transcription activator Sip4.** *EMBO J.* 1999, **18**:6672–6681.

21. Hudson ER, Pan DA, James J, Lucocq JM, Hawley SA, Green KA, Baba O, Terashima T, Hardie DG: **A novel domain in AMP-activated protein kinase causes glycogen storage bodies similar to those seen in hereditary cardiac arrhythmias.** *Curr. Biol.* 2003, **13**:861–866.

22. Polekhina G, Gupta A, Michell BJ, van Denderen B, Murthy S, Feil SC, Jennings IG, Campbell DJ, Witters LA, Parker MW, et al.: **AMPK beta subunit targets metabolic stress sensing to glycogen.** *Curr. Biol.* 2003, **13**:867–871.

23. Gissot L, Polge C, Bouly J-P, Lemaitre T, Kreis M, Thomas M: **AKINbeta3, a plant specific SnRK1 protein, is lacking domains present in yeast and mammals non-catalytic beta-subunits.** *Plant Mol. Biol.* 2004, **56**:747–759.

24. Halford NG, Hey S, Jhurreea D, Laurie S, McKibbin RS, Zhang Y, Paul MJ: **Highly conserved protein kinases involved in the regulation of carbon and amino acid metabolism.** *J. Exp. Bot.* 2004, **55**:35–42.

25. Farrás R, Ferrando A, Jásik J, Kleinow T, Okrész L, Tiburcio A, Salchert K, del Pozo C, Schell J, Koncz C: **SKP1-SnRK protein kinase interactions**

mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.* 2001, **20**:2742–2756.

26. Williams SP, Rangarajan P, Donahue JL, Hess JE, Gillaspy GE: **Regulation of Sucrose non-Fermenting Related Kinase 1 genes in Arabidopsis thaliana.** *Front. Plant Sci.* 2014, **5**:324.

27. Takano M, Kajiya-Kanegae H, Funatsuki H, Kikuchi S: **Rice has two distinct classes of protein kinase genes related to SNF1 of Saccharomyces cerevisiae, which are differently regulated in early seed development.** *Mol. Gen. Genet.* 1998, **260**:388–394.

28. Kanegae H, Miyoshi K, Hirose T, Tsuchimoto S, Mori M, Nagato Y, Takano M: **Expressions of rice sucrose non-fermenting-1 related protein kinase 1 genes are differently regulated during the caryopsis development.** *Plant Physiol. Biochem.* 2005, **43**:669–679.

29. Halford NG, Hardie DG: **SNF1-related protein kinases: global regulators of carbon metabolism in plants?** *Plant Mol. Biol.* 1998, **37**:735–748.

30. Lu C-A, Lin C-C, Lee K-W, Chen J-L, Huang L-F, Ho S-L, Liu H-J, Hsing Y-I, Yu S-M: **The SnRK1A protein kinase plays a key role in sugar signaling during germination and seedling growth of rice.** *Plant Cell* 2007, **19**:2484–2499.

31. Lee K-W, Chen P-W, Lu C-A, Chen S, Ho T-HD, Yu S-M: **Coordinated responses to oxygen and sugar deficiency allow rice seedlings to tolerate flooding.** *Sci. Signal.* 2009, **2**:ra61.

32. Polge C, Thomas M: **SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control?** *Trends Plant Sci.* 2007, **12**:20–28.

33. Rodrigues A, Adamo M, Crozet P, Margalha L, Confraria A, Martinho C, Elias A, Rabissi A, Lumbreras V, González-Guzmán M, et al.: **ABI1 and PP2CA phosphatases are negative regulators of Snf1-related protein kinase1 signaling in Arabidopsis.** *Plant Cell* 2013, **25**:3871–3884.

34. Tsai AY-L, Gazzarrini S: **AKIN10 and FUSCA3 interact to control lateral organ development and phase transitions in Arabidopsis.** *Plant J.* 2012, **69**:809–821.

35. Guérinier T, Millan L, Crozet P, Oury C, Rey F, Valot B, Mathieu C, Vidal J, Hodges M, Thomas M, et al.: **Phosphorylation of p27(KIP1) homologs KRP6 and 7 by SNF1-related protein kinase-1 links plant energy homeostasis and cell proliferation.** *Plant J.* 2013, **75**:515–525.

36. Tiessen A, Prescha K, Branscheid A, Palacios N, McKibbin R, Halford NG, Geigenberger P. **Evidence that SNF1-related kinase and hexokinase are involved in separate sugar-signalling pathways modulating post-translational redox activation of ADP-glucose pyrophosphorylase in potato tubers.** *Plant J.* 2003, **35**:490-500.

37. Radchuk R, Radchuk V, Weschke W, Borisjuk L, Weber H: **Repressing the expression of the SUCROSE NONFERMENTING-1-RELATED PROTEIN KINASE gene in pea embryo causes pleiotropic defects of maturation similar to an abscisic acid-insensitive phenotype.** *Plant Physiol.* 2006, **140**:263–278.

38. Radchuk R, Emery RJN, Weier D, Vigeolas H, Geigenberger P, Lunn JE, Feil R, Weschke W, Weber H: **Sucrose non-fermenting kinase 1 (SnRK1) coordinates metabolic and hormonal signals during pea cotyledon growth and differentiation.** *Plant J.* 2010, **61**:324–338.

39. Zhang Y, Shewry PR, Jones H, Barcelo P, Lazzeri PA, Halford NG: **Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley.** *Plant J.* 2001, **28**:431–441.

40. Townley R, Shapiro L: **Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase.** *Science* 2007, **315**:1726–1729.

41. Amodeo GA, Rudolph MJ, Tong L: **Crystal structure of the heterotrimer core of Saccharomyces cerevisiae AMPK homologue SNF1.** *Nature* 2007, **449**:492–495.

42. Riek U, Scholz R, Konarev P, Rufer A, Suter M, Nazabal A, Ringler P, Chami M, Müller SA, Neumann D, et al.: **Structural properties of AMP-activated protein kinase: dimerization, molecular shape, and changes upon ligand binding.** *J. Biol. Chem.* 2008, **283**:18331–18343.

43. Scholz R, Suter M, Weimann T, Polge C, Konarev P V, Thali RF, Tuerk RD, Viollet B, Wallimann T, Schlattner U, et al.: **Homo-oligomerization and activation of AMP-activated protein kinase are mediated by the kinase domain alphaG-helix.** *J. Biol. Chem.* 2009, **284**:27425–2737.

44. Nayak V, Zhao K, Wyce A, Schwartz M: **Structure and dimerization of the kinase domain from yeast Snf1, a member of the Snf1/AMPK protein family** *Structure* 2006, **14**:477-485.

45. Shen W, Reyes MI, Hanley-Bowdoin L: **Arabidopsis protein kinases GRIK1 and GRIK2 specifically activate SnRK1 by phosphorylating its activation loop.** *Plant Physiol.* 2009, **150**:996–1005.

46. Lee K-W, Chen P-W, Lu C-A, Chen S, Ho T-HD, Yu S-M: **Coordinated responses to oxygen and sugar deficiency allow rice seedlings to tolerate flooding.** *Sci. Signal.* 2009, **2**:ra61.

47. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR: **Abscisic acid: emergence of a core signaling network.** *Annu. Rev. Plant Biol.* 2010, **61**:651–679.

48. Lin C-R, Lee K-W, Chen C-Y, Hong Y-F, Chen J-L, Lu C-A, Chen K-T, Ho T-HD, Yu S-M: **SnRK1A-interacting negative regulators modulate the nutrient starvation signaling sensor SnRK1 in source-sink communication in cereal seedlings under abiotic stress.** *Plant Cell* 2014, **26**:808–827.

49. Crozet P, Margalha L, Confraria A, Rodrigues A, Martinho C, Adamo M, Elias CA, Baena-González E: **Mechanisms of regulation of SNF1/AMPK/SnRK1 protein kinases.** *Front. Plant Sci.* 2014, **5**:190.

50. Ananieva EA, Gillaspy GE, Ely A, Burnette RN, Erickson FL: **Interaction of the WD40 Domain of a Myoinositol Polyphosphate 5-Phosphatase with SnRK1 Links Inositol, Sugar, and Stress Signaling** *Plant Physiol.* 2008, **148**:1868–1882.

51. Bhalerao RP, Salchert K, Bakó L, Okrész L, Szabados L, Muranaka T, Machida Y, Schell J, Koncz C: **Regulatory interaction of PRL1 WD protein with Arabidopsis SNF1-like protein kinases.** *Proc. Natl. Acad. Sci. U. S. A.* 1999, **96**:5322–5327.

52. Lee J-H, Terzaghi W, Gusmaroli G, Charron J-BF, Yoon H-J, Chen H, He YJ, Xiong Y, Deng XW: **Characterization of Arabidopsis and rice DWD proteins and their roles as substrate receptors for CUL4-RING E3 ubiquitin ligases.** *Plant Cell* 2008, **20**:152–167.

53. Sugden C, Crawford RM, Halford NG, Hardie DG: **Regulation of spinach SNF1-related (SnRK1) kinases by protein kinases and phosphatases is associated with phosphorylation of the T loop and is regulated by 5'-AMP.** *Plant J.* 1999, **19**:433–439.

54. Nunes C, Primavesi LF, Patel MK, Martinez-Barajas E, Powers SJ, Sagar R, Fevereiro PS, Davis BG, Paul MJ: **Inhibition of SnRK1 by metabolites: tissue-dependent effects and cooperative inhibition by glucose 1-phosphate in combination with trehalose 6-phosphate.** *Plant Physiol. Biochem.* 2013, **63**:89–98.

55. Ding X, Richter T, Chen M, Fujii H, Seo YS, Xie M, Zheng X, Kanrar S, Stevenson RA, Dardick C, et al.: **A rice kinase-protein interaction map.** *Plant Physiol.* 2009, **149**:1478–1492.

56. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

57. Van Leene J, Witters E, Inzé D, De Jaeger G: **Boosting tandem affinity purification of plant protein complexes.** *Trends Plant Sci.* 2008, **13**:517–520.

58. Jossier M, Bouly J-P, Meimoun P, Arjmand A, Lessard P, Hawley S, Grahame Hardie D, Thomas M: **SnRK1 (SNF1-related kinase 1) has a central role**

in sugar and ABA signalling in Arabidopsis thaliana. *Plant J.* 2009, **59**:316–328.

59. Glinski M, Weckwerth W: **Differential multisite phosphorylation of the trehalose-6-phosphate synthase gene family in Arabidopsis thaliana: a mass spectrometry-based process for multiparallel peptide library phosphorylation analysis.** *Mol. Cell. Proteomics* 2005, **4**:1614–1625.

60. Blanvillain R, Kim JH, Wu S, Lima A, Ow DW: **OXIDATIVE STRESS 3 is a chromatin-associated factor involved in tolerance to heavy metals and oxidative stress.** *Plant J.* 2009, **57**:654–665.

61. Yan Y-S, Chen X-Y, Yang K, Sun Z-X, Fu Y-P, Zhang Y-M, Fang R-X: **Overexpression of an F-box protein gene reduces abiotic stress tolerance and promotes root growth in rice.** *Mol. Plant* 2011, **4**:190–197.

62. Wang X, Peng F, Li M, Yang L, Li G: **Expression of a heterologous SnRK1 in tomato increases carbon assimilation, nitrogen uptake and modifies fruit development.** *J. Plant Physiol.* 2012, **169**:1173–1182.

63 Kato T, Takeda K: **Associations among characters related to yield sink capacity in space-planted rice.** *Crop Sci.* 1996, **36**:1135-1139.

64. Zhao Bu-hong, Wang Peng, Zhang Hong-xi: **Analysis on source-sink and grain-filling characteristics of two-line hybrid rice Yangliangyou 6.** *Rice Science* 2006, **20**:65–72.

65. Jain M, Li Q-B, Chourey PS: **Cloning and expression analyses of sucrose non-fermenting-1-related kinase 1 (SnRK1b) gene during development of sorghum and maize endosperm and its implicated role in sugar-to-starch metabolic transition.** *Physiol. Plant.* 2008, **134**:161–173.

66. Bungard D, Fuerth BJ, Zeng P-Y, Faubert B, Maas NL, Viollet B, Carling D, Thompson CB, Jones RG, Berger SL: **Signaling kinase AMPK activates stress-promoted transcription via histone H2B phosphorylation.** *Science* 2010, **329**:1201–1205.

67. Abate G, Bastonini E, Braun KA, Verdone L, Young ET, Caserta M: **Snf1/AMPK regulates Gcn5 occupancy, H3 acetylation and chromatin remodelling at S. cerevisiae ADY2 promoter** *Biochim. Biophys. Acta - Gene Regul. Mech.* 2012, **1819**:419–427.

68. Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P: **The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid.** *Dev. Cell* 2004, **7**:373–385.

69. O'Brien M, Kaplan-Levy RN, Quon T, Sappl PG, Smyth DR: **PETAL LOSS, a trihelix transcription factor that represses growth in Arabidopsis**

thaliana, binds the energy-sensing SnRK1 kinase AKIN10. *J. Exp. Bot.* 2015, **66**:2475-2485.

70. Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, et al.: **An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes.** *Nat. Protoc.* 2015, **10**:169–187.

71. Katoh K, Standley DM**: MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol. Biol. Evol.* 2013, **30**:772–780.

6.4 Golden2-Like transcription factor: boosting photosynthesis through chloroplast development

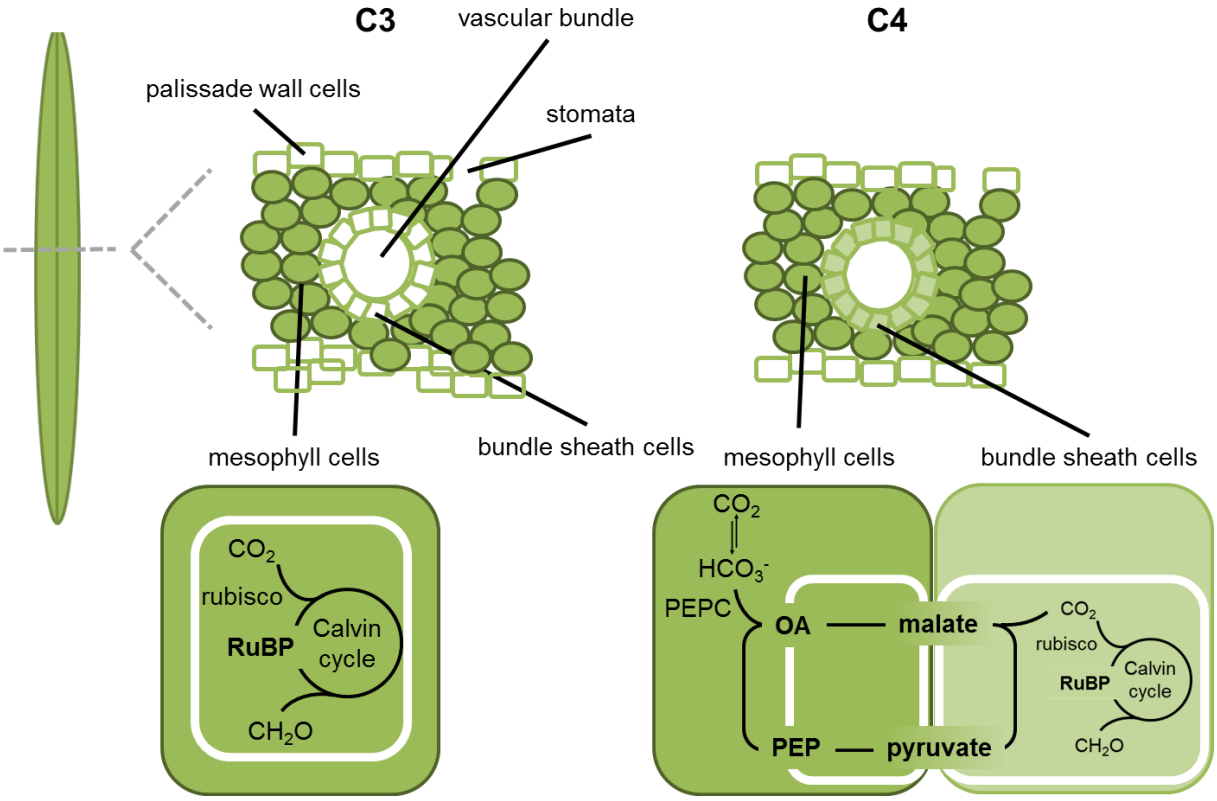
Introduction

Plants use the energy of the sun to 'fix' or convert atmospheric CO_2 into organic compounds. This fixed carbon is used for developing a healthy plant body and to allocate the sugars or starches to the seeds or fruits that we harvest for food. However, not all plants are able to exert this carbon conversion step as efficient as other plants do. The majority of plants, including rice, first fix CO_2 into a three-carbon (C3) compound, phosphoglycerate. This reaction, known as C3 photosynthesis, is catalysed by ribulose bisphosphate carboxylase (RubisCo). This enzyme is inherently inefficient, because it evolved from an environment that initially contained no oxygen. Consequently, RubisCo has low selectivity in distinguishing CO_2 from O_2 as a substrate. Catalysing the reaction with oxygen results in photorespiration instead of photosynthesis, thus dramatically reducing the photosynthetic efficiency. In addition, this selectivity of RubisCo for CO_2 is temperature sensitive, becoming poorer at higher temperatures. Net losses of carbon by this reduced photosynthetic efficiency are estimated at 25% [1].

Some flowering plants evolved to a more efficient photosynthetic pathway with a 4-carbon (C4) compound, oxaloacetate, as the first photosynthetic intermediate. The advantage of C4 over C3 photosynthesis is mainly due to compartmentalisation of reactions, based on the Kranz anatomy of the leaves. In brief, a typical grass leaf is made of a blade and sheath regions delimited by a ligule. The leaf is divided longitudinally by several parallel veins, with the largest vein being the midrib. Surrounding the veins are rings of bundle sheath cells, which are on their turn surrounded by mesophyll cells. In C3 grasses (e.g. wheat, rice, barley) only the mesophyll cells are photosynthetic, whereas in C4 plants (e.g. maize) both bundle sheath and mesophyll cells are photosynthetic (Figure 1). This is the so-called 'Kranz' anatomy. The two cell types in the Kranz anatomy each have different types of chloroplasts that accumulate complementary photosynthetic enzymes. Initially, CO_2 is assimilated in the mesophyll cells by phosphoenol pyruvate carboxylase (PEPC), an enzyme insensitive to oxygen. Malate, derived from oxaloacetate, then diffuses to the bundle sheet cells which contain RubisCo. There, malate is decarboxylated and the released CO_2 is fixed by RubisCo. This mechanism thus concentrates CO_2 around RubisCo in the bundle sheet cells which favours the carboxylation reaction, resulting in higher photosynthetic efficiency [2]. In fact, this also results in a net increase in water and nitrogen use efficiency [2,3]. Because storage of CO_2 in the bundle-sheath cells results in a lowering of the CO_2-levels in the mesophyll cells, CO_2 diffuses more swiftly into the leaf. Stomata can thus be smaller and/or less numerous, preventing water loss due to transpiration. Also, because RubisCo activity is more efficient in C4 plants, less of the protein is needed, saving nitrogen resources [2,3]. C4 plants are therefore better adapted to hotter and dryer environments. Also C4 species that do not rely on the Kranz anatomy were discovered [4,5]. These single cell C4 (SCC4) species also seem to compartmentalise their photosynthetic reactions, but in this case by two types of chloroplasts within the same cells [6].

In both the Kranz anatomy-based as the single cell C4 species, compartmentalisation is only possible with the advent of two complementary types of chloroplasts. A thorough understanding of plastid development, and the formation of dimorphic chloroplasts is therefore crucial to engineer C3 photosynthesis for higher efficiency.

Figure 1. Overview of C3 versus C4 metabolism.



Rice is a C3 photosynthetic organism. This means that only the mesophyll cells are photosynthetic (filled green in the picture), and that all photosynthetic cells contain one type of chloroplasts (outlined in white). These fix CO_2 by RubisCo activity. Due to the ambiguous carboxylase and oxygenase activity of RubisCo, RuBP can be oxidised instead of carboxylated, resulting in loss of substrate and consumption of energy. C4 plants bypass this low selectivity of RubisCo by actively delivering and concentrating CO_2 to a specific compartment where the RubisCo enzyme resides. For this, C4 plants express two types of chloroplasts. One type resides in the mesophyll cells, whereas another type is located in the bundle sheath cells. The chloroplasts in the mesophyll cells incorporate carbon dioxide through the action of PEP carboxylase, an enzyme inert to oxygen. The resulting oxaloacetate is converted to malate and transported to the chloroplasts in bundle sheath cells. There, malate is decarboxylated, and the resulting CO2 is used to generate carbohydrates by the conventional C3 pathway. RubisCo: Ribulose bisphosphate carboxylase; RuBP: Ribulose-1,5-bisphosphate; PEPC: phosphoenol pyruvate carboxylase; OA: oxaloacetate; PEP: phoshpoenol pyruvate.

Chloroplast differentiation in flowering plants is influenced by both environmental and developmental cues. From a developmental perspective, there is a major difference in chloroplast differentiation for C3 compared to C4 plants. C3 plants require only one single chloroplast type for all photosynthetic cells, whereas in C4 plants, chloroplasts will develop from a default state comparable with C3 chloroplasts to specialised forms according to the cell type they reside in [7,8].

Chloroplasts functioning in mesophyll cells will sustain carbon fixation and conversion to malate (by expressing PEPC, pyruvate phosphate dikinase (PPDK) and NADP malate dehydrogenase (NADP-MPDH). Bundle sheath cell chloroplasts on the other hand will accumulate NADP-Malic enzyme (NADP-ME) and RubisCo for malate decarboxylation and actual photosynthesis [9]. Distinct regulatory mechanisms must therefore operate in mesophyll cells and bundle sheath cells of C4 plants to control chloroplast development.

GOLDEN2-like transcription factors

One of the few established transcriptional regulators of chloroplast development are the GOLDEN2-LIKE (GLK) transcription factors. *GOLDEN2* was first characterised in maize [10]. Later it was found that plant species contain one to four paralogous *GOLDEN2*-like genes, and that these regulate chloroplast development [11,12]. The *GLK* genes are members of the GARP (after maize GOLDEN2, the ARR B-class proteins from Arabidopsis, and Chlamydomonas Psr1) superfamily and contain one distant MYB-related domain and a specific GOLDEN2 terminal domain [13,14].

GLK genes are required for chloroplast development and for chloroplast specialization in C4 plants [15]. In maize, two GLKs are expressed; *GOLDEN2* and *GOLDEN2-LIKE 1* (*GLK1*). The *GOLDEN2* gene is expressed in bundle sheath cells whereas the *GLK1* gene is expressed in mesophyll cells [14]. Moreover, the *golden2* mutation led to aberrant bundle sheath chloroplasts. Therefore, the GLK transcription factors could be the genes to trigger functional differentiation of bundle sheath cells. The extent to which compartmentalization of GLK gene function in maize is representative of a more general C4 regulatory mechanism has not yet been investigated however.

GLK transcription factors also occur in pairs in C3 species like *Physcomitrella patens* [15], Arabidopsis [11], tomato [16] and rice [14]. In all cases, the GLK proteins are expressed in all photosynthetic cells. In P. patens and Arabidopsis, the gene pairs are probably redundant in gene function [15]. The rice GLKs show orthologous relationship with their maize counterparts; *OsGLK1* is an ortholog of *ZmGLK1* and *OsGLK2* is an ortholog of *GOLDEN2* [14]. As such, GLK gene duplication in this lineage preceded the speciation of rice and maize and evolution of C4 physiology associated with chloroplast dimorphism [12]. Despite this, *OsGLK1* and *OsGLK2* regulate chloroplast development in both bundle sheath and mesophyll cells [12].

Studies from different C3 plant species pointed out that GLKs are involved in light adaptation in plants. Observations from rice overexpression lines for *OsGLK1* strongly suggest that *OsGLK1* is controlled by light and phytohormones, and that it is a key regulator of chloroplast development [17]. A potential pathway for OsGLK1 is that the transcription factor upregulates sigma factor genes (SIG) and other nuclear encoded genes for photosynthetic machinery. Then SIGs, together with PEP (plastid-encoded plastid RNA polymerase) core subunits, induce expression of plastid encoded genes for photosynthetic machinery. Consequently, proplastids develop into chloroplasts and acquire photosynthetic function [17]. Experiments from Arabidopsis showed that AtGLK1 and AtGLK2 interact with

proline-rich regions of G-box-binding bZIP factors [18]. GLK1 is functionally redundant with GLK2. Double mutants are pale green in all photosynthetic tissues and show reduced granal thylakoids in chloroplasts [11]. This pale green phenotype was also observed for rice double mutants [19].

In tomato, GLK1 was found to be more important in leaves while GLK2 is more predominant in fruits. Overexpression of *GLKs* resulted in increased chlorophyll and carotenoid content, as well as increased photosynthesis and carbohydrate accumulation [16]. *Physcomitrella patens* GLK genes regulate chloroplast development in the moss [15]. Together, these findings indicate that transcription factors encoded by GLK genes positively regulate chloroplast development by a mechanism conserved widely in the plant kingdom.

Golden2-like transcription factor in TraitMill

The rice *GOLDEN2*-like gene *OsGLK1* was screened in TraitMill. Whereas its paralog *OsGLK2* contains a nuclear localisation signal (NLS) similar to GOLDEN2 in maize and the Arabidopsis GLKs, GLK1 has no classical NLS [14]. The gene was overexpressed under control of PRO_{GOS2}, a medium constitutive promoter, and showed a positive yield phenotype expressed as total weight of the seeds per plant. This yield phenotype was mainly driven by a higher number of seeds combined with a concomitant higher fill rate as compared to null segregants. Fill rate is an indication of filling of the seeds and is defined as the proportion of number of filled seeds over the number or florets.

Tandem affinity purification of OsGLK1

We screened the OsGLK1 protein as bait in tandem affinity purification experiments, followed by mass spectrometry (TAP-MS) to identify potential interactors. For this, we designed bait constructs encompassing an N- and C-terminal fusion of the GS tag, driven by the PRO_{35S}, and introduced these constructs in rice callus. We performed two replicate experiments from 50mg total protein on transgenic callus expressing each fusion construct. These four independent TAP experiments from callus resulted in the identification of four interactors that could be confirmed from at least two independent purifications (Figure 2). Additionally, we found one protein that was retrieved in only one experiment. Remarkably, we only found the bait and an importin for the N-terminal fusion, suggesting that this construct might not be functional.

Figure 2. Interaction data for the OsGLK1 bait.

	accession	description	symbol	PRO_{35S}:GS-GLK1	PRO_{35S}:GLK1-GS	total
bait	loc_os06g24070	GOLDEN2-LIKE 1	OsGLK1	2	2	4
light signalling related	loc_os01g62060	plant-specific domain TIGR01589 family protein			2	2
	loc_os02g53140	CONSTITUTIVE PHOTOMORPHOGENIC 1	COP1		1	1
importins	loc_os04g26841	IMPORTIN 8	IMP8	2	2	4
	loc_os04g26850	SUPER SENSITIVE TO ABA AND DROUGHT 2	SAD2		2	2
	loc_os02g27470	IMPORTIN 7	IMP7		2	2

Interaction table for proteins retrieved from TAP experiments from rice T0 callus. For both N- and C-terminal fusions, two replicate purifications were performed on 50mg of total protein extract. Potential interactors that were retrieved in multiple bait groups and normally would have been assigned as non-specific/false positive are marked red.

GLK1 interactors from rice callus mainly confirm role in light perception

From our interaction dataset, we found two proteins potentially related to light perception and signalling. One of these is annotated as the rice ortholog of *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*). COP1 is a protein involved in ubiquitin-mediated protein degradation and represses photomorphogenesis [20], in part by degradation of transcription factors which mediate light response [21].

The other is plant-specific domain TIGR01589 family protein. This interactor has a very similar expression pattern to *OsGLK1*, as could be derived from Genevestigator data for anatomy and development (Figure S1). The upstream sequence of the gene contains a G-box, a well-characterized light-responsive element (LRE) [22]. However, presence of the G-box alone in a gene's promoter is insufficient to confer light regulation of gene expression [23]. Data from the two potential Arabidopsis orthologs at3g55240 (68.42% coverage, 80% similarity) and at5g02580 (70.33% coverage, 79.69% similarity) further hinted towards involvement in the light signalling pathway. Both genes also contain a core G-box in their promoter regions. For at3g55240, binding of the LONG HYPOCOTYL 5 (HY5) transcription factor to its promoter was confirmed experimentally [24]. HY5 is a basic leucine zipper (bZIP) transcription factor involved in light-regulated transcriptional activation of G-box-containing promoters [25]. In addition, the gene was found to contain an unstable, circadian clock-regulated transcript [26]. Plants overexpressing the gene were phenotypically characterised as pale green, with fast plant development and taller growth than wild-type plants [27]. This phenotype resembles that of wild type plants grown under weak light conditions. Therefore the gene was designated *PEL* for '*PSEUDO-ETIOLATION IN LIGHT*' [27]. The *PEL* overexpressors had normal photosynthetic activity and relatively normal chloroplast structure except for a lower level of membrane stacking and starch accumulation [27]. Transcripts of the gene were preferentially expressed in rosette leaves in wild-type plants. Downregulation of the gene by RNAi led to plant death

at very early stages of development, hence the knockout phenotype of the at4g55240 gene may be lethal [27]. At5g02580 is induced by PHYTOCHROME INTERACTING FACTOR 5 (PIF5), and its promoter region is bound by PIF1/PIL5 and PIF3 [28–30]. Phytochrome interacting factors (PIFs) are in general involved in regulating response in absence to light. They repress chlorophyll and carotenoid biosynthesis by binding sequence-specifically to a core DNA G-box motif [31].

From these observations from Arabidopsis, it is likely that the TIGR01589 family protein indeed is involved in chloroplast development. The etiolated phenotype resulting from overexpressing one of the homologs suggests that it might be an antagonist of OsGLK1. A poplar homolog [63.37% coverage, 76.56% similarity] was tested in TraitMill under control of the PRO_{GOS2} and also showed a negative trend in Greenness before flowering. This fits with the phenotype from PEL overexpression in Arabidopsis.

Conclusion

Corn GOLDEN2, was one of the first transcription factors identified through genetic analysis as potentially playing a role in C4 photosynthesis [32]. Also, the *GOLDEN2* transcript was enriched in bundle sheath cells, and GLK1 was enriched in mesophyll cells [10,14]. The function of *GOLDEN2* and its paralog *GLK1* were further elucidated through the characterisation of the homologs in Arabidopsis thaliana. Double mutants of the two Arabidopsis *GLK* genes remained pale green throughout development and showed reduced thylakoid stacking [11], a phenotype that was also observed for rice double mutants [19]. Furthermore, *AtGLK1* and *AtGLK2* were found functionally redundant. Recent studies demonstrated that the regulation of chloroplast development by light, auxin/cytokinin and plastid-derived retrograde signals is dependent on the GOLDEN2-LIKE transcription factors (GLKs), which are required for the expression of several chlorophyll biosynthesis genes [34], [35]. Characterization of transcriptional targets of AtGLK1 and AtGLK2 identified at least 100 targets, most of which encode components of the photosynthetic apparatus, such as photosystem I and II, as well as tetrapyrrole biosynthesis [33]. The function of GLK transcription factors seems to be conserved in rice, as GLK1 controls the expression of photosynthesis-related nuclear genes [17]. Overexpression of *GLK1* induces chloroplast development in rice. Therefore, GLK transcription factors are positive regulators of chloroplast development.

Overexpression of GLK transcription factors can be beneficial for fruit/seed quality. This was shown in tomato, where overexpression of GLKs increased fruit quality without affecting non-fruit parts. This means that overexpression of GLKs can be employed as a means of affecting fruit quality through improved sugar levels, carotenoids, organic acids which also translates to favourable flavours. They also found that quality effects of the transcription factors are additive, which presents a potential for amplified fruit quality improvement [16]. Related to that, in Arabidopsis only *AtGLK2* was expressed in siliques, and mutants for this gene showed pale siliques [11].

We screened the chloroplast development regulator OsGLK1 for potential interactors in rice callus and co-purified two potential light signalling regulators of which one is novel. Of these two interactors, COP1 is the best established. COP1

is an E3 ubiquitin ligase that represses photomorphogenesis [20] in part by degradation of transcription factors which mediate light responses. In that light, GLK1 could fit as a potential novel COP1 target in rice.

Another interactor we found is a TIGR01589 family protein. We deduced potential involvement in chloroplast development, as the phenotypes of Arabidopsis overexpression lines and rice plants overexpressing a *Populus trichocarpa* homolog showed a paler green phenotype. This is in line with the phenotypes of *COP1* overexpression lines and *glk* mutants which in fact also show an etiolated phenotype in Arabidopsis [11,21]. We therefore suggest that the TIGR01589 family protein and COP1 have a similar role as negative regulators of chloroplast development, antagonizing OsGLK1.

To have a clear view on the exact functioning of the TIGR01589 family protein in the chloroplast development pathway, additional experiments would be required. For example, TAP experiments under normal light conditions, specifically from leaf samples could point out the context-specific relevance of the interactions we detected. Further, it would be interesting to see if the dark-grown phenotype of TIGR01589 family protein overexpression lines resemble the *cop1* phenotype (skotomorphogenesis in the dark).

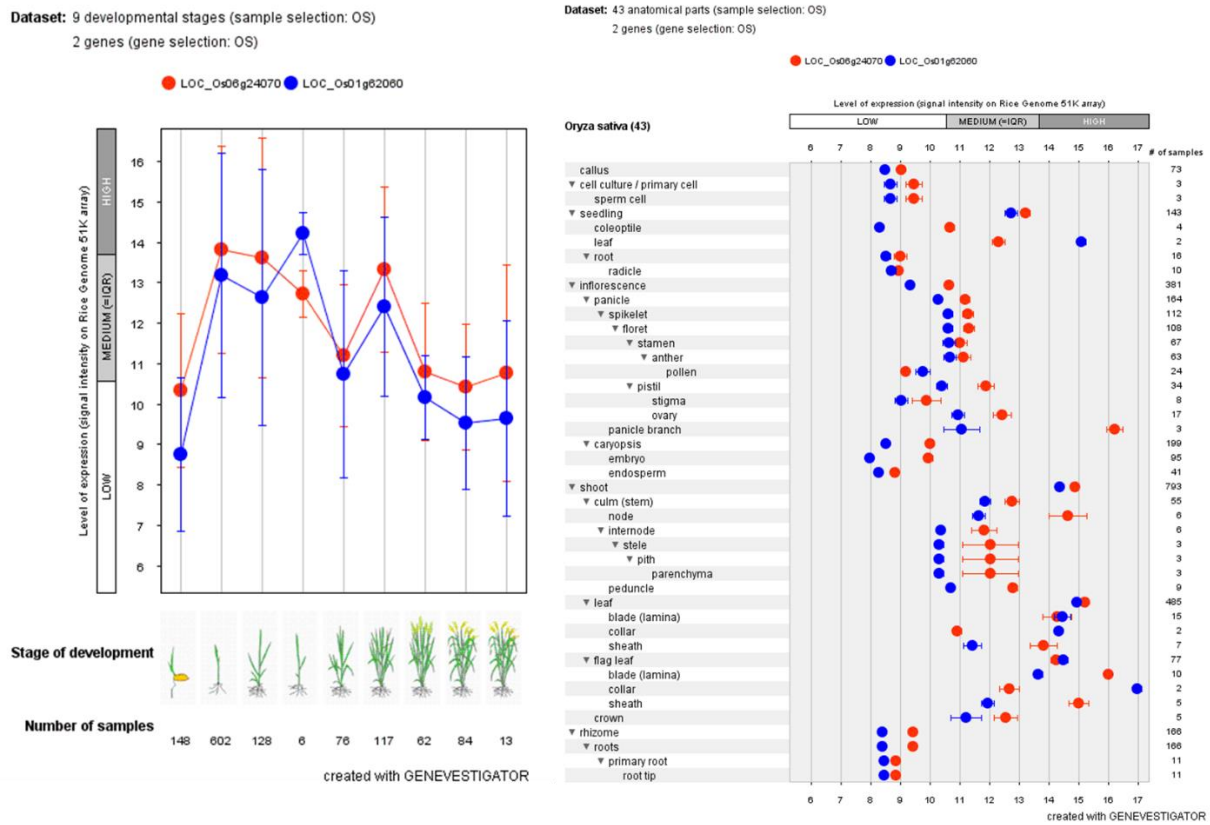
We conclude that our TAP approach at least paved the way to uncover new players in connecting light signalling with chloroplast development.

Materials & Methods

Screening of the OsGLK1 bait in the rice TAP platforms
Screening of the bait in callus, T0 shoots and T1 seedlings was done as described in the materials and methods in chapter 6.1.

Supplementary information

Figure S1. *OsGLK1* and the *PEL*-homolog display similar expression profiles in rice.

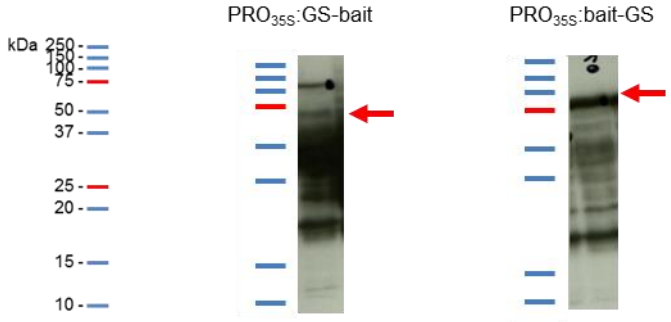


Data retrieved from Genevestigator [34] for OsGLK1 (loc_os06g24070) and the PEL-homolog (loc_os01g62060) for development (left panel) and anatomy (right panel).

Figure S2. Primers used in this chapter.

primer n°	name	primer sequence
prm27565	attB1 kozak ATG OsGLK1	ggggacaagtttgtacaaaaaagcaggctccaccatgcttgccgtg
prm27566	attB2 no stop OsGLK1	ggggaccactttgtacaagaaagctgggtctccacacgctgg
prm27567	attB2r no start OsGLK1	ggggacagctttcttgtacaaagtggccatgcttgccgtgtcg
prm27568	attB3 stop OsGLK1	ggggacaactttgtataataaagttggtcatccacacgct

Figure S3. Expression analysis of the OsGLK1 bait protein in callus.



Protein extracts of the transgenic lines were analyzed by immunoblotting with peroxidase anti-peroxidase antibody against the GS tag to verify protein levels of the OsGLK1 bait N- (left) and C- (right) terminal tagged with the GS tag. Molecular marker values are provided in the left panel in kilo Dalton (kDa). The molecular weight of the tagged proteins is 74.6 kDa.

Supplementary background list and mass spec files.

Supplementary files can be found through the following link:

https://floppy.psb.ugent.be/public.php?service=files&t=fb92196c73b5e9eaed39e3678dd62d78 (password: rice_TAP).

The mass spec files for SnRK1 interaction data described in this chapter can be found in SI_Chapter_6_4_GLK1_MS_data.xlsx.

Author contribution

The PhD candidate was in charge of generating the TAP constructs, the transformations, maintenance and upscaling of callus cell lines and plant material, analysis of the mass spectrometry data. He also wrote the manuscript. TAP purifications and mass spectrometry measurements were mainly done by employees of VIB.

References

1. Foyer CH, Bloom AJ, Queval G, Noctor G: **Photorespiratory metabolism: genes, mutants, energetics, and redox signaling** *Annu. Rev. Plant Biol.* 2009*,* **60**: 455-484.

2. Raghavendra A, Sage R: **C4 photosynthesis and related CO2 concentrating mechanisms.** *Adv. in Photosynthesis and Respiration.* 2010.

3. Leegood RC: **Strategies for engineering C(4) photosynthesis.** *J. Plant Physiol.* 2013, **170**:378–388.

4. Voznesenskaya E V, Franceschi VR, Kiirats O, Freitag H, Edwards GE: **Kranz anatomy is not essential for terrestrial C4 plant photosynthesis.** *Nature* 2001, **414**:543–546.

5. Voznesenskaya E V., Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE: **Proof of C4 photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae).** *Plant J.* 2002, **31**:649–662.

6. Sharpe RM, Offermann S: **One decade after the discovery of single-cell C4 species in terrestrial plants: what did we learn about the minimal requirements of C4 photosynthesis?** *Photosynth. Res.* 2014, **119**:169–180.

7. Langdale JA, Nelson T: **Spatial regulation of photosynthetic development in C4 plants.** *Trends Genet.* 1991, **7**:191–196.

8. Nelson T, Dengler N: **Leaf Vascular Pattern Formation.** *Plant Cell* 1997, **9**:1121–1135.

9. Langdale JA: **C4 cycles: past, present, and future research on C4 photosynthesis.** *Plant Cell* 2011, **23**:3879–3892.

10. Hall LN, Rossini L, Cribb L, Langdale JA: **GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf.** *Plant Cell* 1998, **10**:925–936.

11. Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA: **GLK gene pairs regulate chloroplast development in diverse plant species.** *Plant J.* 2002, **31**:713–727.

12. Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al.: **Evolution of GOLDEN2-LIKE gene function in C(3) and C (4) plants.** *Planta* 2013, **237**:481–495.

13. Riechmann JL, Heard J, Martin G, Reuber L, -Z. C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al.: **Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes** *Science* 2000, **290**:2105–2110.

14. Rossini L, Cribb L, Martin DJ, Langdale JA: **The maize golden2 gene defines a novel class of transcriptional regulators in plants.** *Plant Cell* 2001, **13**:1231–1244.

15. Yasumura Y, Moylan EC, Langdale JA: **A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants.** *Plant Cell* 2005, **17**:1894–1907.

16. Powell ALT, Nguyen C V, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons C, Fernández-Muñoz R, Vicente A, et al.: **Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development.** *Science* 2012, **336**:1711–1715.

17. Nakamura H, Muramatsu M, Hakata M, Ueno O, Nagamura Y, Hirochika H, Takano M, Ichikawa H: **Ectopic overexpression of the transcription factor OsGLK1 induces chloroplast development in non-green rice cells.** *Plant Cell Physiol.* 2009, **50**:1933–1949.

18. Tamai H, Iwabuchi M, Meshi T: **Arabidopsis GARP transcriptional activators interact with the Pro-rich activation domain shared by G-box-binding bZIP factors.** *Plant Cell Physiol.* 2002, **43**:99–107.

19. Waters MT, Langdale JA: **The making of a chloroplast.** *EMBO J.* 2009, **28**:2861–2873.

20. Lau OS, Deng XW: **The photomorphogenic repressors COP1 and DET1: 20 years later.** *Trends Plant Sci.* 2012, **17**:584–593.

21. McNellis TW, von Arnim AG, Deng XW: **Overexpression of Arabidopsis COP1 results in partial suppression of light-mediated development: evidence for a light-inactivable repressor of photomorphogenesis.** *Plant Cell* 1994, **6**:1391–1400.

22. Donald RG, Cashmore AR: **Mutation of either G box or I box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter.** *EMBO J.* 1990, **9**:1717–1726.

23. Puente P, Wei N, Deng XW: **Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis.** *EMBO J.* 1996, **15**:3732–3743.

24. Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y, Tongprasit W, Zhao H, Lee I, Deng XW: **Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development.** *Plant Cell* 2007, **19**:731–749.

25. Chattopadhyay S: **Arabidopsis bZIP Protein HY5 Directly Interacts with Light-Responsive Promoters in Mediating Light Control of Gene Expression** *Plant Cell* 1998, **10**:673–684.

26. Gutierrez RA, Ewing RM, Cherry JM, Green PJ: **Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: rapid decay is associated with a group of touch- and specific clock-controlled genes.** *Proc. Natl. Acad. Sci. U. S. A.* 2002, **99**:11513–11518.

27. Ichikawa T, Nakazawa M, Kawashima M, Iizumi H, Kuroda H, Kondou Y, Tsuhara Y, Suzuki K, Ishikawa A, Seki M, et al.: **The FOX hunting system: an alternative gain-of-function gene hunting technique.** *Plant J.* 2006, **48**:974–985.

28. Oh E, Kang H, Yamaguchi S, Park J, Lee D, Kamiya Y, Choi G: **Genome-wide analysis of genes targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during seed germination in Arabidopsis.** *Plant Cell* 2009, **21**:403–419.

29. Leivar P, Tepperman JM, Cohn MM, Monte E, Al-Sady B, Erickson E, Quail PH: **Dynamic Antagonism between Phytochromes and PIF Family Basic Helix-Loop-Helix Factors Induces Selective Reciprocal Responses to Light and Shade in a Rapidly Responsive Transcriptional Network in Arabidopsis** *Plant Cell* 2012, **24**:1398–1419.

30. Zhang Y, Mayba O, Pfeiffer A, Shi H, Tepperman JM, Speed TP, Quail PH: **A Quartet of PIF bHLH Factors Provides a Transcriptionally Centered Signaling Hub That Regulates Seedling Morphogenesis through Differential Expression-Patterning of Shared Target Genes in Arabidopsis** *PLoS Genet.* 2013, **9**:e1003244.

31. Leivar P, Quail PH: **PIFs: pivotal components in a cellular signaling hub.** *Trends Plant Sci.* 2011, **16**:19–28.

32. Langdale JA, Kidner CA: **BUNDLE-SHEATH DEFECTIVE, a mutation that disrupts cellular differantiation in maize leaves.** *Development* 1994, **120**:673–681.

33. Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA: **GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis.** *Plant Cell* 2009, **21**:1109–1128.

34. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P: **Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes.** *Adv. Bioinformatics* 2008, **2008**: 420747.

Part V. General conclusions and perspectives

Chapter 7. Discussion and perspectives

For our food supply, we depend almost exclusively on plants either directly or indirectly. Increasing crop yield is therefore a major challenge for modern agriculture. One of the major food crops, rice, feeds more than half of the world's population. Further, it is a model organism for other cereals as corn and wheat. A lot of research efforts were made to genetically engineer crops to transform them into more stress resilient varieties and/or varieties that provide more seed yield. To achieve this, it will be necessary to understand how the molecular machinery that determines these parameters operate. To date, merely individual genes regulating components of plant growth or seed yield are known, but their connections are largely unexplored. Nevertheless, biological processes are often carried out by the dynamic interaction of proteins in complexes and signalling pathways. Unravelling these interconnections can thus greatly speed up our understanding of complex traits as yield. In that light, the main objective of this work was to build a platform for high-throughput screening of protein-protein interactions in rice.

Establishment of a TAP-MS platform in rice

To achieve this goal, we optimised a Tandem Affinity Purification coupled to Mass Spectrometry (TAP-MS) protocol in rice, and subsequently used the technology to study the protein-protein interactions of genes with a potential yield effect. Currently, TAP is the most powerful method to isolate protein complexes and discover unknown protein associations. Yeast-two hybrid (Y2H) and protein complementation assay (PCA) also allow for screening of protein-protein interactions, but rather provide a binary read-out. They are therefore complementary to TAP-MS [1]. Other PPI screening methods such as protein arrays, which rely on immobilization of proteins on chemically derivatized glass slides to enable interrogation of interaction partners [2], and size-exclusion chromatography coupled to mass spectrometry (SEC-MS) look promising but are still in their infancy for use in plant protein interaction studies [2–4].

We established a TAP platform in rice by optimizing each of the different building blocks – promoter, TAP tag, expression vector - required for making a TAP construct, streamlining every step in the AP-MS workflow, implementing the latest and most sensitive MS technology, and further integrating recent technical advances from research in Arabidopsis. In parallel with fine-tuning the protocol, we assayed its quality by screening interaction partners for CKS1, APC10 and CDKD. These three baits are proteins that participate in well-known, conserved complexes. The improvements we implemented allowed to apply the protocol on a portfolio of different types of rice biomass, ranging from tissues that provide a wealth of protein extract such as cultured cells, but also more technically demanding tissues, such as whole plants, or even specifically isolated organs or tissues.

With the tools for an optimised protocol at hand, we screened multiple baits related to enhanced growth or seed yield in the different rice tissues available. We assayed in total 28 baits in callus cells, and succeeded in recovering 25 of the 28 baits in a reproducible manner. For 24 of the 25 recovered baits we retrieved interacting proteins in at least two independent purifications. This amounts to 85.7% of the baits screened. Also for plants, we obtained similar success rates. Six of the in total seven baits tested in T1 seedlings were recovered, and for all six at least one interactor was retrieved reproducibly. The numbers from experiments on T0 shoots are harder to compare, as we performed nonreplicate experiments. Here, nine out of the ten baits were detected, with eight of them co-purifying at least one interactor in one experiment. Summarized, we obtained significantly higher success rates with our protocol as compared to the rice kinase study of Rohila *et al.* [5,6] (23%). We are aware that kinases are a tricky class of baits to screen interactors from, and that screening these kinases by using only N-terminal tagging might have compromised the 'true' success rate of the their protocol. On the other hand, they calculated the success rate from nonreplicate experiments. We screened in total 5 kinases (including CDKD), and were able to reproducibly retrieve interactors for all of them.

In this work, three baits related to enhanced growth or seed yield were highlighted. TA_HLH is related to the PRE/ILI family of HLH proteins and resulted in an increased seed size phenotype when overexpressed. SnRK1 is an evolutionary conserved fuel gauge, controlling plant growth and development in response to carbon availability. OsGLK1 is involved in chloroplast development and regulation and gave a yield increase phenotype, expressed as total seed weight per plant. We will further discuss different aspects of our TAP workflow, with references to the baits highlighted in this work to pinpoint potential room for improvement or for variations on the standard procedure that could be envisioned for future work.

With our TAP workflow, we enabled to make a choice in the biological material to screen for interactors. Each biomass type has its merits, and we recommend careful consideration of the pro's and con's before selecting a biomass type to screen a protein of interest in. Using callus cells as biological starting material allows to screen complexes in reasonable timelines with high success rates. Moreover, callus cells allow for easy upscaling if higher levels of protein input would be required. Currently we proliferate the callus on gel medium. This in contrast to cell suspension cells, which are grown in liquid medium and allow for easy addition or removal of chemicals or hormones. A possible extension of our protocol could therefore be to transfer transformed microcalli into liquid medium to start a cell suspension culture [7]. This system would then allow to track complex dynamics in response to addition or removal of chemicals or hormones from the liquid medium [8–10]. It would for example be interesting to assay the complex constitution for SnRK1 in sucrose-depleted conditions, as the kinase is active under these conditions. An obvious limitation in using cultured cells arises when the gene encoding the protein of interest has a discrete expression profile in either developmental or anatomical context. Also, in our current protocol, callus cells are

indeed not greened, and grown under continuous light conditions. This could result in false negatives when the bait is involved in a biological process that is linked to circadian clock or light regulation. For example, the screening of GLK1 in callus cells grown in continuous light might not have been ideal, as GLK1 is involved on chloroplast development. It will be interesting to see if GLK1 screened from plant or leaf tissues would provide a more specific protein interaction profile.

To allow screening of PPI related to these more specialised pathways, we extended our protocol to plants. For this, we established a short-term protocol based on T0 shoot and a longer one based on 2 weeks old T1 seedlings. These can provide a more comprehensive developmental picture of complexes. In the SnRK1 interaction data for example, we retrieved more potential kinase targets related to metabolism as sucrose phosphate synthases, multiple trehalose synthase isoforms and a bifunctional aspartokinase-homoserine dehydrogenase from purifications from plant tissues as compared to callus cells. As shoots or seedlings are composed from a mixture of organs, complex dynamics in relation to organ development is lost. In principle, this could partially be solved by mapping expression data on the protein interaction data. Regulation at the post-transcriptional level is in that case not accounted for however. For discretely expressed baits, sensitivity of the protocol can also drop, as there is a dilution effect of the relevant tissues with the remainder of the plant's cells.

We anticipate that assaying specific tissues will gain more and more momentum in the interactomics field. First, sampling specifically the tissues where a bait is functioning increases sensitivity of the protocol while maintaining the context in plant development. We purified complexes for APC10 and CDKD from the proliferation zone of the emerging 4th leaf as a proof of concept. Complex isolation from minute samples opens possibilities for elucidating biological processes by comparing protein complexes assayed from different organs or from organs at different developmental stages. For this, rice is a more suitable model compared to Arabidopsis, since sampling from its larger organs is less problematic for this type of experiments. It will be interesting to test other tissue types as caryopsis or inflorescence.

We opted for an overexpression strategy to drive TAP construct expression in wild type background. Our study made use of two constitutive promoters, PRO_{GOS2}, the rice *GOS2* promoter, and PRO_{35S}, the CaMV 35S promoter. We are aware that overexpression might lead to aberrant levels of the bait that exceed normal physiological conditions. This can result in unspecific associations lowering the signal to noise. The optimal situation to preserve conditions implies completely replacing the endogenous protein by the tagged version, similar to the result after homologous recombination. A possible surrogate for homologous recombination could be CRISPR-mediated insertion of affinity tags [11–14], but to our knowledge this was not yet tested in plants. Alternatively an additional copy of the tagged transgene under control of the endogenous locus could be introduced through BAC transgenesis [15]. We are however interested in the biological context of the overexpression constructs, as these were responsible for the increased yield

phenotype. In that sense, experiments with PRO_{GOS2}-driven expression could gain momentum as the *GOS2* promoter is most frequently used for TraitMill screening.

For isolation of the complexes, we preferred developing a two-step affinity purification protocol as this provides a higher signal-to-noise ratio. Indeed, identification of protein complex components requires that they are present in sufficient absolute as well as relative amounts. While the Orbitrap-MS mass spectrometer is extremely sensitive, the more limited dynamic range and loading capacity of the MS and LC-column respectively may hamper detection of *bona fide* complex components among an excess of non-specific contaminant proteins. Possibly, that explains why testing a single-affinity protocol using GFP did not yield better results in our hands (data not shown). It is worth noting however that improvements in sample handling practices and available reagents have largely superseded the need for tandem affinity procedures to obtain high signal/low background results, especially in the yeast and mammalian field. There, single-step affinity capture has proven sufficient, and being shorter in duration, it increases the chances of observing labile interactors [16,17].

A major challenge associated with TAP-MS experiments is making the distinction between *bona fide* interactions that make biological sense and non-specific background interactions. In this work, we used frequency filtering based on the principle of re-occurrence of contaminating proteins to address this issue. This approach integrates information from all runs, i.e. 174 purifications on 38 different bait proteins, which allows to zoom in on the specificity of different protein associations. Briefly, for evaluating the interactors of a specific bait protein our filtering method uses all the purifications of other baits as negative control purifications. As baits belonging to the same process can show overlapping interaction profiles, we rather considered bait groups. These were assigned dependent on biological relatedness. With more purifications on different baits performed in the future, the threshold currently set will need to be re-evaluated. The background list will thus evolve according to the status of the TAP platform. An important remark here is that this way of filtering discriminates specific from non-specific interactors rather than *bona fide* from background identifications. This became apparent in purifications from the SnRK1 and TA_HLH baits. For the SnRK1 interaction dataset, we were able to revoke some proteins from the background as 'true' interactors because their interaction with the SnRK1 complex was previously described. Similarly, in the TA_HLH interaction dataset, we assigned UV DAMAGED DNA-BINDING 1A – normally considered background – as genuine interactor as Y2H experiments in Arabidopsis showed physical interaction with another genuine interactor DDB1 AND CULLIN4 ASSOCIATED FACTOR1 [18].

Therefore, quantitative data from the mass spectrometer can be implemented for evaluating each interactor. Since true interactors are expected to be enriched throughout purification compared to a control, the enrichment should be reflected in a higher measure of quantity. In addition, the quantitative aspect enables determining the changes in the dynamics in protein complex composition during development.

A myriad of possible methods exist for relative quantification. In brief, isotope labelling methods based on either metabolic or chemical labelling are not frequently used in plants because of insufficient uptake for incorporation of the labels or because of issues with specificity [19], respectively. Alternatively label-free MS methods are generally considered. Here, computational approaches are used to obtain quantitative information from MS data. Label-free quantification is less accurate compared to labelling strategies, so reliable quantification requires a higher number of replicates with an additional statistical or probabilistic analysis. A nice overview of the existing methods is provided in [20]. Summarized, two types of quantitative mass spectrometry methods exist, based on either analysis of the MS1 spectra of the MS2 spectra. Identification of proteins by MS usually involves two measurement steps of peptides generated by a site-specific enzyme. In a first 'survey' scan (MS1) the m/s of the intact ionized peptide is monitored. A pre-set amount of most abundant peptides are then isolated and selected for fragmentation. The fragments' m/z ratios are recorded in the second 'fragment' spectrum (MS2). These fragment ion spectra are then assigned to peptide sequences and ultimately proteins using computational approaches for peptide. In general, methods based on MS1 intensity are potentially more accurate than methods based on MS2 spectra. During shotgun sequencing only the most intense MS1 peptides are selected for MS2. In other words, there is a bias of MS2 data towards high abundance peptides, which limits the accuracy of measurements in the low abundance range. The drawback of using MS1 data is that these rely on alignment of the peaks with the retention time of the LC. Therefore, the purifications and LC runs need to be carried out as reproducible as possible [21].

Even after subtraction of background proteins, the read-out of TAP experiments is still merely a list of proteins that co-precipitated with the bait. This implies that the configuration of direct interactions between these proteins is not reflected, it is even not clear whether one or more (different) complexes are retrieved. Further, our current platform only provides a static snapshot of the protein complexes under investigation, whereas in reality many proteins belong to complexes with a dynamically changing composition depending on the general state of the biological process they function in.

Quantitative mass spectrometry can apart from discriminating background binders, also be used for charting the dynamics of protein complexes. With the platform we developed, quantitative analysis can be applied to assay composition of complexes during organ development by sampling tissues at different time points in growth. Another option is the mapping of orthogonal data layers onto the detected interactions. For example, interactors with strong expression correlation could hint to a common molecular assembly [22]. The expression correlation can be deduced directly, from protein expression datasets, or indirectly though co-expression data. This was illustrated for the *PSEUDO ETIOLATION IN LIGHT* (*PEL*)-homolog found in the GLK1 dataset, which showed a very similar transcription expression profile as the *GLK1* bait. In support, protein sub-localisation data provide supportive information towards the co-localisation of interaction partners

in a specific cellular compartments. In a similar way, annotation data as gene ontology or MapMan annotation can indicate whether interaction partners are functioning in common biological pathways or exerting a similar molecular function.

Another missing layer of information is the directionality of the protein links. In most cases, proteins influence the functioning of their interaction partner in a positive or a negative way. A study in Drosophila computationally integrated PPI networks with genetic RNAi screens to predict the directions of the interactions [23]. They were able to construct a 'signed' PPI network identifying positive and negative regulators of signalling pathways and protein complexes. In that sense, phenotypic data can greatly aid in assessing the directionality of the interactions. For example, the overexpression phenotype of *PEL*, the Arabidopsis homolog of an interactor found in the GLK1 dataset, was pale green. This is contrary to the phenotype of GLK1 overexpression, suggesting that GLK1 and PEL might be antagonists in chloroplast development. Whether PEL is acting up-or downstream of GLK1 still needs to be clarified.

Added value of interaction data resulting from the TAP workflow

This work can be considered as a 'model example' of how a certain cutting-edge technology was developed at an academic research institute and then – once the technology perfected – has been transferred to the biotech industry. The established TAP workflow in rice enables to obtain a 'molecular snapshot' around a specific protein of interest. For any biotech company, this type of information is of enormous value, as it provides a unique piece of additional information to help understanding the molecular mechanisms that steer a lead gene's yield enhancement phenotype. This also provides an advantage towards competitors, as to our knowledge only Syngenta probably has a similar platform [24]. We envision that the interaction data resulting from our TAP workflow can further be employed as both a hypothesis-generating tool and as a framework for data integration and network analysis. Both reason from the 'guilt by association' concept [25]. Within a complex, each protein may have a specialized function that contributes to the overall function of the complex [26]. Thus, proteins participating in a common complex might provide the same yield effect or result in additional or synergistic effects when combined with the initial lead. Interaction data can also help in unravelling the molecular function of a lead, as 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.

The success of the TraitMill phenotyping platform greatly depends on the genes that are selected as input. Until recently a single transgene approach was used to test the effect in rice plants. The unique database of genes that showed a significant positive effect on growth or seed yield enables CropDesign to unravel yet uncharted networks that may steer growth or seed yield. A better view in the underlying molecular mechanisms will allow to make more targeted gene choices for TraitMill screening which will raise the lead hit ratio. In addition, the field is

moving towards multi transgene approaches to further increase yield enhancing effects. This strategy implies testing the effect of two potential lead genes in the same plant, and one expects that two genes from a same or two interacting pathways could deliver additive or even synergistic effects.

Orthogonal data used to gain confidence in the interactors such as co-expression data, text mining and/or phenotypic data can be further utilised to prioritise which interactor to select for phenotypic screening for a yield-enhancing effect. The same datasets can also provide hints for designing the transgenic construct. For example, if interactors lead to a similar phenotype when mutated or overexpressed, there is a good chance that they co-operate within the complex. Interactors that rather show contrasting phenotypes on the other hand, will probably antagonise within the complex. For example, overexpression of the ANAPHASE-PROMOTING COMPLEX 10 (APC10) subunit of the anaphase promoting complex (APC) results in larger leaves [27], whereas the plant-specific negative regulator of the APC, SAMBA, needs to be downregulated to result in the same phenotype [28]. Expression data can provide additional hints in which promoter to choose to screen the new yield-enhancement candidate with.

The resulting candidate can be selected for screening either as a single gene construct, or stacked with another interactor or the initial lead. When stacking two genes, the epistatic effect aimed for is in fact the result of genetic interaction between the two stacking partners. Genetic interactions reflect functional relationships between genes, in which the phenotypic effect of one gene is modified by another. Genetic interactions are generally assayed by comparing the effect of mutating each gene individually to the effect of the double mutant. In the extreme, synthetic lethality can occur when the combination of two mutations leads to lethality [29]. In plants, synthetic lethal genetic interactions have been explored mainly by RNAi to detect genes whose products act in the same essential pathway [30,31]. In yeast, large networks of genetic interactions have been measured genome-wide [32]. Analysis of these datasets learned that 10-20% of the protein-protein interactions were also genetic interactions, which is significantly higher than expected randomly (~3%) [32]. Moreover, PPI found from TAP-MS showed the highest enrichment in overlap (20%) compared to PPI found from Y2H (10%) and PCA (~10%) assays [32]. Conversely however, only a small fraction of the genetic interaction pairs (~1%) were also physically linked [32]. This suggests that the vast majority of genetic interactions occurs between rather than within protein complexes.

These findings were derived from screening of mutants for synthetic lethality in single cells. Therefore it is difficult to assess whether these will still hold true for data from TraitMill, which is generally derived from overexpression (and mostly gain of function) screening from a multicellular organism. We anticipate that the more straightforward ways to derive stacking constructs from TAP data would probably be through combining the components that are limiting within a complex or by integrating factors that provide links between different molecular assemblies. The former case has been illustrated through the epistatic effects from combining GIF and GRF5 for leaf size [33], bZIP10 and ABI3 or bZIP25 and ABI3 for the expression of a reporter gene for seed specific expression [34] and TSI and TSIP

for enhanced pathogen resistance and salt tolerance [35]. These examples are mainly an illustration of combining both partners of heterodimers, such as transcription factor complexes, each required for transcription activation/repression. For the latter case, it is important to notice that Y2H interactions rather tends to deliver links between pathways, whereas TAP data tends to unravel the molecular assemblies [36].

Protein-protein interaction data also provide an excellent framework for further data integration and subsequent network analysis. There are different types of molecular networks that have been developed and associational networks are the most frequently used. These networks –akin to social networks as Facebook or LinkedIn – are based on the principle that one can guess things about a gene (or person) based on other genes (or people) it is connected to. For example, properties of genes can be identified from omics data and used to link the genes that share the properties, resulting in a co-function network [37]. Co-function networks can be really powerful when used in the context of specific biological contexts.

Perspectives

Proteins do not only interact with other proteins in a cellular environment. Transcription factors target specific binding elements on DNA to recruit or block the transcription machinery. Traditionally, chromatin immunoprecipitation (ChIP) is used for the identification of genomic regions where specific proteins are associated. An antibody is required to enrich for the protein bound to the DNA locus. The associated region is then identified by qPCR or sequencing. The use of antibodies can be omitted by tagging the DNA binding protein by an affinity tag, and performing the appropriate affinity purification protocol (chromatin affinity purification or ChAP). This was for example done for studying the binding of VERNALIZATION (VRN) proteins to the FLOWERING LOCUS C (FLC) using TAP-tagged VRN2 and VRN5. The association of VRN proteins with the FLC locus was quantified by qPCR after enrichment of cross-linked VRN baits by single step IgG purification [38]. More recently a tandem chromatin affinity purification or TChAP protocol was proposed [39]. By using a double step enrichment using the cross-linking-resistant His-Biotin-His-tag, the DNA enrichment ratios were improved compared to ChIP and ChAP. Combined with Illumina sequencing, it enables the identification of novel TF-DNA interactions. The downside is that more material is required. Therefore the method was currently only used in cell suspension cultures [40,41]. A TChAP approach can also be applied for studying protein-RNA interactions. This was for example done by Zhu et al, who used a combination of tandem affinity purification with Illumina sequencing to study the interplay of 2 ARGONAUTE proteins with miRNA's in regulating shoot apical meristem [42].

Instead of the protein-centric approach for analysing protein-DNA interactions, also the 'reverse', a DNA-centric approach, is possible. Here, a DNA sequence is used to pull down proteins that bind to it. In this manner, transcription factors and other DNA binding proteins that bind to a specific DNA region can be identified.

Reverse-ChIP is based on adding heterologous cis elements to a genomic region of interest. Cells are transformed with both the heterologous cis-elements flanking the genomic region and a tagged heterologous DNA binding protein which binds to the cis-elements [43]. Prior to purification, *in vivo* crosslinking is used to maintain all interactions. The complex, including the genomic region of interest and its DNA binding proteins, are then affinity purified. After elution and de-crosslinked, the components are identified by mass spectrometry and/or qPCR. The approach was recently further modified, by using affinity-tagged TALE (for transcription activator-like effector) and CRISPR (for clustered regularly interspaced short palindromic repeats) molecules respectively that target a specific DNA region of interest [44,45]

Another variation using epitope tagged proteins was used for the study of protein-lipid interactions for yeast proteins [46]. This shows that establishing an AP-MS protocol is worthwhile to study protein-protein interactions and can further be extended in exploring interactions of the bait protein with all other kinds of biomolecules.

In conclusion, we established a tandem affinity purification coupled to mass spectrometry protocol for *Oryza sativa*. This method can be employed for the screening of protein-protein interaction of specific proteins of interest. Information on the protein complex can help in the analysis of biological networks, functional gene analysis and gene discovery. In this work, we used the method to elucidate the protein interaction context for genes that are related to increased growth or seed yield. This will help in the functional characterization of these specific genes and their relation to their yield enhancing characteristics. However, in the nearby future, this technique can also be adapted to study other types of interactions, as protein-DNA, protein-RNA or protein-lipid interactions. Hence, by developing this protocol of TAP in rice, we have extended the molecular toolbox of this model species and helped to pave a way for its continued success as model species for cereal crops.

References

1. Yu H, Braun P, Yildirim M a, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al.: **High-quality binary protein interaction map of the yeast interactome network.** *Science* 2008, **322**:104–110.

2. Popescu SC, Popescu G V, Bachan S, Zhang Z, Seay M, Gerstein M, Snyder M, Dinesh-Kumar SP: **Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays.** *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:4730–5.

3. Popescu SC, Popescu G V, Bachan S, Zhang Z, Gerstein M, Snyder M, Dinesh-Kumar SP: **MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays.** *Genes Dev.* 2009, **23**:80–92.

4. Aryal UK, Xiong Y, McBride Z, Kihara D, Xie J, Hall MC, Szymanski DB: **A Proteomic Strategy for Global Analysis of Plant Protein Complexes** *Plant Cell* 2014, **26**:3867-3882.

5. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

6. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

7. Nallamilli BRR, Zhang J, Mujahid H, Malone BM, Bridges SM, Peng Z: **Polycomb group gene OsFIE2 regulates rice (Oryza sativa) seed development and grain filling via a mechanism distinct from Arabidopsis.** *PLoS Genet.* 2013, **9**:e1003322.

8. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Pérez AC, Chico JM, Bossche R Vanden, Sewell J, Gil E, et al.: **NINJA connects the co-repressor TOPLESS to jasmonate signalling.** *Nature* 2010, **464**:788–791.

9. Bassard J-E, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, et al.: **Protein-protein and protein-membrane associations in the lignin pathway.** *Plant Cell* 2012, **24**:4465–4482.

10. Antoni R, Gonzalez-Guzman M, Rodriguez L, Peirats-Llobet M, Pizzio G a, Fernandez M a, De Winne N, De Jaeger G, Dietrich D, Bennett MJ, et al.: **PYRABACTIN RESISTANCE1-LIKE8 plays an important role for the regulation of abscisic acid signaling in root.** *Plant Physiol.* 2013, **161**:931–941.

11. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al.: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339**:819–823.

12. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339**:823–826.

13. Park A, Won ST, Pentecost M, Bartkowski W, Lee B: **CRISPR/Cas9 allows efficient and complete knock-in of a destabilization domain-tagged essential protein in a human cell line, allowing rapid knockdown of protein function.** *PLoS One* 2014, **9**:e95101.

14. Böttcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, Wieland I, Gaul U, Förstemann K: **Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells.** *Nucleic Acids Res.* 2014, **42**:e89.

15. Zhou R, Benavente LM, Stepanova AN, Alonso JM: **A recombineering-based gene tagging system for Arabidopsis.** *Plant J.* 2011, **66**:712–723.

16. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin Z-Y, Breitkreutz B-J, Stark C, Liu G, et al.: **A global protein kinase and phosphatase interaction network in yeast.** *Science* 2010, **328**:1043–1046.

17. Oeffinger M: **Two steps forward--one step back: advances in affinity purification mass spectrometry of macromolecular complexes.** *Proteomics* 2012, **12**:1591–1608.

18. Zhang Y, Feng S, Chen F, Chen H, Wang J, McCall C, Xiong Y, Deng XW: **Arabidopsis DDB1-CUL4 ASSOCIATED FACTOR1 forms a nuclear E3 ubiquitin ligase with DDB1 and CUL4 that is involved in multiple plant developmental processes.** *Plant Cell* 2008, **20**:1437–1455.

19. Karp NA, Huber W, Sadowski PG, Charles PD, Hester S V, Lilley KS: **Addressing accuracy and precision issues in iTRAQ quantitation.** *Mol. Cell. Proteomics* 2010, **9**:1885–1897.

20. Pardo M, Choudhary JS: **Assignment of protein interactions from affinity purification/mass spectrometry data.** *J. Proteome Res.* 2012, **11**:1462–1474.

21. Hubner NC, Mann M: **Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC).** *Methods* 2011, **53**:453–459.

22. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, et al.: **Predictive models of molecular**

machines involved in Caenorhabditis elegans early embryogenesis. *Nature* 2005, **7052**:861-865.

23. Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumüller RA, Mohr SE, Perrimon N: **Integrating protein-protein interaction networks with phenotypes reveals signs of interactions.** *Nat. Methods* 2014, **11**:94-99.

24. Zhong J, Haynes PA, Zhang S, Yang X, Andon NL, Eckert D, Yates JR, Wang X, Budworth P: **Development of a System for the Study of Protein - Protein Interactions in Planta : Characterization of a TATA-Box Binding Protein Complex in Oryza sativa.** *Journal of Proteome Research.* 2003, **2**: 514-522.

25. Drewes G, Bouwmeester T: **Global approaches to protein–protein interactions.** *Curr. Opin. Cell Biol.* 2003, **15**:199–205.

26. Bauer A, Kuster B: **Affinity purification-mass spectrometry.** *Eur. J. Biochem.* 2003, **270**:570–578.

27. Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GTS, Inzé D, Ferreira PCG: **The APC/C subunit 10 plays an essential role in cell proliferation during leaf development.** *Plant J.* 2011, **68**:351–363.

28. Eloy NB, Gonzalez N, Van Leene J, Maleux K, Vanhaeren H, De Milde L, Dhondt S, Vercruysse L, Witters E, Mercier R, et al.: **SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:13853–13858.

29. Beyer A, Bandyopadhyay S, Ideker T: **Integrating physical and genetic maps: from genomes to interaction networks.** *Nat. Rev. Genet.* 2007, **8**:699–710.

30. Hartung F, Suer S, Bergmann T, Puchta H: **The role of AtMUS81 in DNA repair and its genetic interaction with the helicase AtRecQ4A.** *Nucleic Acids Res.* 2006, **34**:4438–4448.

31. Yin X-J, Volk S, Ljung K, Mehlmer N, Dolezal K, Ditengou F, Hanano S, Davis SJ, Schmelzer E, Sandberg G, et al.: **Ubiquitin lysine 63 chain forming ligases regulate apical dominance in Arabidopsis.** *Plant Cell* 2007, **19**:1898–1911.

32. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, et al.: **The genetic landscape of a cell.** *Science* 2010, **327**:425–431.

33. Horiguchi G, Kim G-T, Tsukaya H: **The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana.** *Plant J.* 2005, **43**:68–78.

34. Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J: **Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2.** *J. Biol. Chem.* 2003, **278**:21003–21011.

35. Ham B-K, Park JM, Lee S-B, Kim MJ, Lee I-J, Kim K-J, Kwon CS, Paek K-H: **Tobacco Tsip1, a DnaJ-type Zn finger protein, is recruited to and potentiates Tsi1-mediated transcriptional activation.** *Plant Cell* 2006, **18**:2005–2020.

36. Yu H, Braun P, Yildirim M a, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al.: **High-quality binary protein interaction map of the yeast interactome network.** *Science* 2008, **322**:104–110.

37. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY: **Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana.** *Nat. Biotechnol.* 2010, **28**:149–156.

38. De Lucia F, Crevillen P, Jones AME, Greb T, Dean C: **A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:16831–16836.

39. Verkest A, Abeel T, Heyndrickx KS, Van Leene J, Lanz C, Van De Slijke E, De Winne N, Eeckhout D, Persiau G, Van Breusegem F, et al.: **A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification.** *Plant Physiol.* 2014, **164**:1122–1133.

40. Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han S-K, Jégu T, Archacki R, Van Leene J, Andriankaja M, et al.: **ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development.** *Plant Cell* 2014, **26**:210–229.

41. Heyman J, Cools T, Vandenbussche F, Heyndrickx KS, Van Leene J, Vercauteren I, Vanderauwera S, Vandepoele K, De Jaeger G, Van Der Straeten D, et al.: **ERF115 controls root quiescent center cell division and stem cell replenishment.** *Science* 2013, **342**:860–863.

42. Zhu H, Hu F, Wang R, Zhou X, Sze S-H, Liou LW, Barefoot A, Dickman M, Zhang X: **Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development.** *Cell* 2011, **145**:242–256.

43. Byrum SD, Raman A, Taverna SD, Tackett AJ: **ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus.** *Cell Rep.* 2012, **2**:198–205.

44. Byrum SD, Taverna SD, Tackett AJ: **Purification of a specific native genomic locus for proteomic analysis.** *Nucleic Acids Res.* 2013, **41**:e195.

45. Fujita T, Fujii H: **Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR.** *Biochem. Biophys. Res. Commun.* 2013, **439**:132–136.

46. Gallego O, Betts MJ, Gvozdenovic-Jeremic J, Maeda K, Matetzki C, Aguilar-Gurrieri C, Beltran-Alvarez P, Bonn S, Fernández-Tornero C, Jensen LJ, et al.: **A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae.** *Mol. Syst. Biol.* 2010, **6**:430.

Part VI. Summary

Chapter 8. Executive summary

Summary

Agriculture needs to continuously improve to sustain growing food demand of an increasing world population. Further, to preserve the planet's sustainability in terms of flora and fauna, global agricultural productivity should increase without significantly increasing the cultivated area. In other words, most of the increase in production will have to be achieved by increasing the yield per area. This poses an enormous pressure on the agricultural industry to produce more output than ever before [1]. The improvement of intrinsic yield qualities in crops is therefore one of the main goals in the agro-industry. Conventional breeding provided significant improvements, but is limited within the boundaries of the species. The use of biotechnology tools enables to break that genetic boundary, but also poses huge challenges. To decide which promoter-gene combination could alter a plant's intrinsic yield characteristics, thorough knowledge is required about where and when growth regulators are crucial in contributing to an enhanced yield phenotype.

In the past decades, this was tackled by bluntly empirical testing the effect of changing expression levels of genes in model plants and extrapolating the results to crops. This is illustrated by two private initiatives. The company Mendel biotech for example tested numerous transcription factors by constitutive overexpression in Arabidopsis. CropDesign developed TraitMill, an automated plant evaluation platform allowing high-throughput testing of the effect of transgenes on yield and other agronomical valuable traits in rice. Rice is considered as an appropriate model for cereal crops such as maize and wheat with its rather small genome (389 Mb) and evolutionary relatedness. These efforts led to the elucidation of numerous individual growth regulators, but knowledge from these scattered data is still far from sufficient to efficiently engineer complex biological traits such as growth or seed yield. Finding the links between individual growth regulators will be key to determine which molecular networks play in defining yield traits. A significant part of connections between biological entities is defined by protein-protein interactions (PPI). PPI are indeed known to form the basis of many cellular processes and biological functions. Elucidation of the molecular interactions between yield stimulating proteins, collectively called the 'yield interactome', could therefore gain insight in how molecular networks control complex phenotypes.

The main objective of this work was therefore to build a platform for high-throughput screening of protein-protein interactions in rice. For this, we optimised a tandem affinity purification coupled to mass spectrometry (TAP-MS) protocol in this excellent model for cereal crops.

We established a TAP platform in rice by optimizing each of the different building blocks – promoter, TAP tag, expression vector - required for making a TAP construct, streamlining every step in the AP-MS workflow, implementing the latest and most sensitive MS technology, and further integrating recent technical advances from research in Arabidopsis. In parallel with fine-tuning the protocol,

we assayed its quality by screening interaction partners for CKS1, APC10 and CDKD. These three baits are proteins that participate in well-known, conserved complexes. The improvements we implemented allowed to apply the protocol on a portfolio of different types of rice biomass, ranging from tissues that provide a wealth of protein extract such as cultured cells, but also more technically demanding tissues, such as whole plants, or even specifically isolated organs or tissues. Complex isolation from minute samples opens possibilities for elucidating biological processes by comparing protein complexes assayed from different organs or from organs at different developmental stages.

With the tools for an optimised protocol at hand, we screened multiple baits related to enhanced growth or seed yield in the different rice tissues available. We assayed in total 28 bait in callus cells, 25 of which we recovered in a reproducible manner. Of these 25 recovered baits, 24 had interacting proteins that were reproducibly found. This amounts to 85.7% of the baits screened. A total of seven baits were tested in T1 seedlings. Here, the numbers boil down to six of the seven baits recovered reproducibly, and for all six, at least one interactor was retrieved reproducibly. Hence, the success rate is 85.7% for T1 seedlings. Summarized, we obtained significantly higher success rates with our protocol as compared to the rice kinase study of Rohila *et al.* [2,3].

In this work, three baits related to enhanced growth or seed yield were highlighted. TA_HLH is related to the PRE/ILI family of HLH proteins and resulted in an increased seed size phenotype when overexpressed. SnRK1 is an evolutionary conserved fuel gauge, controlling plant growth and development in response to carbon availability. OsGLK1 is involved in chloroplast development and regulation and gave a yield increase phenotype, expressed as total seed weight per plant. In all cases, the TAP read-out provided a unique insight in the working mechanisms of these baits. For TA_HLH, we found diverse subfamilies of helix-loop-helix proteins, which hint towards the existence of a trimodular regulation mechanisms for cell elongation in rice, comparable as what was found in Arabidopsis. For SnRK1 we found both interactors that might influence the activity of the kinase and potential kinase substrates. And last, two interactors for OsGLK1 are potentially involved in light signalling and might be antagonising OsGLK1 function in chloroplast development.

In conclusion, we established a tandem affinity purification coupled to mass spectrometry protocol for *Oryza sativa*. This method can be employed for the screening of protein-protein interaction of specific proteins of interest. Information on the protein complex can help in the analysis of biological networks, functional gene analysis and gene discovery. In this work, we used the method to elucidate the protein interaction context for genes that are related to increased growth or seed yield. This will help in the functional characterization of these specific genes and their relation to their yield enhancing characteristics. However, in the nearby future, this technique can also be adapted to study other types of interactions, as protein-DNA, protein-RNA or protein-lipid interactions. Hence, by developing this protocol of TAP in rice, we have extended the molecular toolbox of this model

species and helped to pave a way for its continued success as model species for cereal crops.

References

1. Alexandratos N, Bruinsma J: **World agriculture towards 2030/2050: the 2012 revision**. *ESA Work. Pap.* 2012, **No. 12-03**:1–154.

2. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

3. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

Samenvatting

De landbouwsector blijft innoveren om een steeds groeiende vraag naar voedsel van een toenemende wereldbevolking te ondersteunen. Daarnaast wordt rekening gehouden met duurzaamheid in termen van zowel flora als fauna. Dat betekent dat de globale landbouwproductiviteit in principe moet toenemen zonder het totaal oppervlak aan landbouwgrond te vergroten. In andere woorden, het merendeel van de productie zal afhangen van een verhoogde opbrengst per oppervlakte [1]. Dit legt een enorme druk op de agrarische industrie om meer output dan ooit te produceren. Het verbeteren van opbrengsten in gewassen is daarom een van de hoofddoelen van de agro-industrie. Conventioneel veredelen biedt significante verbetering, maar is gelimiteerd binnen de grenzen van species. Het gebruik van biotechnologie staat toe om die 'genetische grens' te verbreken, maar stelt ons tegelijk voor gigantische uitdagingen. Om te beslissen welke combinatie van promoter en gen precies nodig is om de opbrengst in een gewas te verhogen, is een degelijke kennis nodig over waar en wanneer groei regulatoren cruciaal zijn in het bijdragen aan dat phenotype.

In de voorbije tien jaar werd dit probleem voornamelijk opgelost door het bluntweg empirisch testen van veranderingen in expressieniveaus van genen in model planten, en de resultaten daarna te extrapoleren naar gewasssen. Twee private initiatieven zijn hier een schoolvoorbeeld van. Het bedrijf Mendel Biotech testte talrijke transcriptiefactoren via constitutieve overexpressie in de zandraket. CropDesign ontwikkelde TraitMill, een geautomatiseerd plant evaluatie platform dat toestaat om het effect van transgenen in rijst in hoge doorvoer te testen op agronomisch waardevolle kenmerken. Rijst wordt hier als gepast model voor graangewassen als maïs en tarwe beschouwd, gezien het evolutionair gerelateerd is en een redelijk klein genoom (398 Mb) heeft. Deze initiatieven leidden tot de opheldering van talrijke individuele groeiregulatoren, maar kennis uit deze verspreide data is ver van voldoende om efficient complexe biologische kenmerken als groei of zaadopbrengst om te bouwen. Het vinden van linken tussen de verscheidene groeiregulatoren zal daarom één van de sleutels zijn in het definiëren van welke moleculare netwerken meespelen in het bepalen van opbrengst. Een signifcant deel van de connecties tussen biologische entiteiten wordt gedefinieerd door eiwit-eiwit interacties. Van deze is immers geweten dat ze de basis vormen van veel cellulaire processen en biologische functies. Het ophelderen van de molecularie interacties tussen opbrengst verhogende eiwtten, 'het opbrengst interactoom', zou daarom inzicht kunnen bieden in hoe moleculaire netwerken complexe phenotypes controlleren.

Het hoofddoel van dit werk is het bouwen van een platform dat instaat voor een hoge-doorvoer screening van eiwit-eiwit-interacties in rijst. Hiertoe optimaliseerden we een 'tandem affinity purification coupled to mass spectrometry' (TAP-MS) protocol in dit uitstekend model voor graangewassen.

We slaagden in het opbouwen van een TAP-MS platform in rijst door elke bouwsteen nodig voor het maken van een TAP construct te optimaliseren, elke stap binnen de workflow te stroomlijnen, de laatste en meest gevoelig MS technologie in te bouwen, en de meest recente technische vooruitgang uit

onderzoek in Arabidopsis te integreren. Parallel met het verfijnen van het protocol, schatten we de kwaliteit in door te screenen naar interactiepartners voor CKS1, APC10 en CDKD. Deze 3 baits zijn eiwitten die deel uitmaken van goed gekende en geconserveerde complexen. De verbeteringen die we aanbrachten stelden ons in staat om het TAP protocol toe te passen op een portfolio van verschillende types rijst biomassa. Deze varieren van weefsels die een weelde aan eiwitextract bieden, zoals callus cellen, tot meer technisch veeleisende weefsels zoals planten en zelfs specifiek geisoleerde organen of stukjes weefsel. Het isoleren van complexen uit minuscule stalen opent mogelijkheden tot het ophelderen van biologische processen door vergelijking van complexen uit verschillende organen of uit organen in een verschillend ontwikkelingsstadium.

Met dit geoptimaliseerd protocol voorhanden, screenden we verscheidene baits die gerelateerd zijn aan groei of verhoogde opbrengst in de verschillende beschikbare types biomassa. We testten in totaal 28 baits in callus cellen. Hiervan werden er 25 op een reproduceerbare wijze teruggevonden. Voor 24 van deze 25 baits, pikten we ook interactoren op in op zijn minst twee verschillende zuiveringen. Dit komt neer op 85,7% van de geteste baits. Ook voor planten bekwamen we eenzelfde 'success rate'. Zes van de in totaal 7 baits gescreend in T1 zaailingen konden we succesvol terugvinden, en deze zes leidden allen tot reproduceerbare identificatie van interactoren. Samengevat verkregen we duidelijk hogere success rates in vergelijking met de rijst kinases studie van Rohila *et al.* [2,3].

In deze thesis werden 3 baits gerelateerd aan verhoogde groei en zaadopbrengst naar voor gebracht. TA_HLH is verwant aan de PRE/ILI familie van helix-loop-helix eiwitten en overexpressie leidde tot grotere zaden. SnRK1 is een evolutionair geconserveerde 'brandstofmeter' die plantengroei en -ontwikkeling controlleert als reactie op beschikbaarheid van koolstof. OsGLK1 is betrokken in de ontwikkeling en regulatie van chloroplasten. Overexpressie van OsGLK1 gaf verhoogde opbrengst, uitgedrukt als totaal gewicht van de zaden per plant. In alle gevallen gaven de TAP resultaten een uniek inzicht in de werkingsmechanismen van deze baits. Voor TA_HLH vonden we verscheidene subfamilies helix-loop-helix eiwitten, die aanwijzing geven naar het bestaan van een trimodulair regulatiemechanisme voor cel elongatie in rijst, vergelijkbaar met Arabidopsis. Voor SnRK1 vonden we zowel interactoren die waarschijnlijk de activiteit van het kinase beïnvloedden als enkele mogelijke kinase substraten. Twee interactoren voor OsGLK1 tenslotte, zijn waarschijnlijk mee betrokken in licht signalisatie en zouden OsGLK1 tegenwerken als regulator van chloroplast ontwikkeling.

Samengevat hebben we een TAP-MS protocol ontwikkeld voor *Oryza sativa.* De methode kan gebruikt worden voor het screenen van eiwit-eiwit interacties voor bepaalde eiwitten van interesse. De resulterende informatie rond het eiwitcomplex kan helpen in het analyseren van biologische netwerken, in het ontrafelen van de genfunctie en in het ontdekken van nieuwe genen. Wij hebben de methode aangewend om de eiwit interactie context voor genen die gerelateerd zijn aan groei of verhoogde zaadopbrengst op te helderen. Dit zal helpen in het functioneel karakteriseren van deze genen en hun relatie met opbrengst verhoging. De

bouwstenen voor het TAP platform kunnen in de nabije toekomst ook aangewend worden om andere types interacties te bestuderen, zoals eiwit-DNA, eiwit-RNA of eiwit-lipide interacties. Door de ontwikkeling van dit TAP-MS platform in rijst hebben we dus de moleculaire gereedschapskist tot het verbeteren van graangewassen helpen uitbreiden.

References

1. Alexandratos N, Bruinsma J: **World agriculture towards 2030/2050: the 2012 revision**. *ESA Work. Pap.* 2012, **No. 12-03**:1–154.

2. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice.** *Plant J.* 2006, **46**:1–13.

3. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, et al.: **Protein-protein interactions of tandem affinity purified protein kinases from rice.** *PLoS One* 2009, **4**:e6685.

Thanks

With about finishing my manuscript and (finally!) getting to the expression of my gratitude to the many people I'd like to thank, I needed to reflect on how it all started. I was finishing my master thesis, and was looking for someone who could help me in setting up a Baekeland project I really wanted to do. Without Geert and his group, and the work they had done before this would never have been possible. I think that in this case the statement of 'standing on the shoulders of giants' is quite suitable. So thank you Geert, Jelle, Nancy, Dominique, Geert P, Jan, Leen, Bernard, Astrid, Aurine and in a later phase Nienke, Michiel and Caroline. Not only for your hard work of course, but also for the help, the good advices, the kindness. Although I was not that much around, it still felt familiar with you guys.

Also thank you so much Steven! At the time I don't think I realised how much effort it must have been to convince the 'guys above' within BASF to approve this project. Therefore I was primarily happy when you said you had good and bad new (the bad news was that I had to start writing a project proposal), but now I am also very grateful.

I think I need to thank Geert and Steven again by the way, for their support and superb guidance throughout my PhD as my promoters. I enjoyed every one of our update meetings and it was always reassuring to have two big-shots of knowledge behind my back. Later on, we were re-enforced by Durba and Srijeet. To both of you also thank you so much for your good advices and help.

In the meantime, I feel myself very lucky being adopted in the Sysbio group. Tom, thanks for sharing the non-confidential room with me. It was really a pleasure! I am sure that the Germans will soon uncover your talents. Annelies, it is good to know that you are passing your talented genes on to the next generation. Berenice, you are a great "mother hen" of Sysbio! Koen, thank you for letting me join this fantastic group.

Also thanks to each of my jury members for your time and precious suggestions to improve the quality of this manuscript.

Ook mijn vrienden en familie mogen niet ontbreken natuurlijk! Simon, Tom, cici, p-man, Helder en Jellyman: ik heb prachtige tijden beleefd in de maagden- en akkerstraat. Het zijn momenten die ik voor altijd zal koesteren. Jullie zijn vrienden voor het leven. O's, Donny en Zlatan, weet dat ik jullie graag zie. Jullie zijn elk op jullie manier kanjers! Hoyo, ik denk niet dat er iets is dat ik niet aan je te danken heb…Bedankt voor je zorg en liefde, je bent de allerbeste mama! En je introduceerde Wilfried, deze topper moet je houden! Luc, Kaat, Lieselot, Geert, Rovertje: dankjewel om mij op te nemen in jullie leuke familie!

Hannelore, we wisten het eigenlijk al véél vroeger, maar ik ben blij dat we er toch achter gekomen zijn dat we voor elkaar gemaakt zijn. Bedankt om me zo goed te soigneren tijdens het schrijven en om van ons stulpje ons paradijs te maken. Ik hou zielsveel van je.