Large-Scale Arabidopsis Phosphoproteome Profiling Reveals Novel Chloroplast Kinase Substrates and Phosphorylation Networks[1][W]

Sonja Reiland, Gaëlle Messerli, Katja Baerenfaller, Bertran Gerrits, Anne Endler, Jonas Grossmann, Wilhelm Gruissem, and Sacha Baginsky*

Department of Biology, Eidgenössische Technische Hochschule Zurich, 8092 Zurich, Switzerland (S.R., G.M., K.B., B.G., A.E., J.G., W.G., S.B.); Functional Genomics Center Zurich, 8057 Zurich, Switzerland (B.G., J.G.); and Institute of Plant Biology, University of Zurich, 8008 Zurich, Switzerland (A.E.)

We have characterized the phosphoproteome of Arabidopsis (*Arabidopsis thaliana*) seedlings using high-accuracy mass spectrometry and report the identification of 1,429 phosphoproteins and 3,029 unique phosphopeptides. Among these, 174 proteins were chloroplast phosphoproteins. Motif-X (motif extractor) analysis of the phosphorylation sites in chloroplast proteins identified four significantly enriched kinase motifs, which include casein kinase II (CKII) and proline-directed kinase motifs, as well as two new motifs at the carboxyl terminus of ribosomal proteins. Using the phosphorylation motifs as a footprint for the activity of a specific kinase class, we connected the phosphoproteins with their putative kinases and constructed a chloroplast CKII phosphorylation network. The network topology suggests that CKII is a central regulator of different chloroplast functions. To provide insights into the dynamic regulation of protein phosphorylation, we analyzed the phosphoproteome at the end of day and end of night. The results revealed only minor changes in chloroplast kinase activities and phosphorylation site utilization. A notable exception was ATP synthase β-subunit, which is found phosphorylated at CKII phosphorylation sites preferentially in the dark. We propose that ATP synthase is regulated in cooperation with 14-3-3 proteins by CKII-mediated phosphorylation of ATP synthase β-subunit in the dark.

Protein phosphorylation is an important posttranslational modification in eukaryotic cells that regulates many cellular processes. Recent technical and conceptual advances in phosphopeptide enrichment strategies and the improvement of mass spectrometric analysis instruments have facilitated the large-scale comprehensive analysis of protein phosphorylation (for review, see Preisinger et al., 2008; Baginsky, 2009). Phosphopeptides are usually enriched by affinity chromatography prior to mass spectrometric analysis in order to minimize ion suppression effects from nonphosphorylated peptides (Salih, 2005). Common phosphopeptide enrichment strategies use immobilized metal affinity (IMAC) or titanium dioxide (TiO_2) affinity chromatography. Both methods deliver largely complementary peptide identifications, suggesting that a combination of different methods is necessary to achieve comprehensive phosphoproteome coverage (Bodenmiller et al., 2007). Examples of plant phosphoproteome analyses include the plasma membrane of cultured Arabidopsis (*Arabidopsis thaliana*) cells and the leaf (Nuhse et al., 2004, 2007; de la Fuente van Bentem et al., 2006; Benschop et al., 2007). The most comprehensive phosphopeptide profiling from cultured Arabidopsis cells revealed 2,597 phosphopeptides from 1,346 phosphoproteins (Sugiyama et al., 2008). All phosphoproteome data are available in the PhosphAT database (http://phosphat.mpimp-golm.mpg.de/).

Current plant phosphoproteome data show that protein phosphorylation occurs in all subcellular compartments. The nucleus accounts for more than 40% of all detected phosphoproteins in some data sets (de la Fuente van Bentem et al., 2006; Sugiyama et al., 2008). Gene Ontology (GO) classification of phosphoproteins revealed an overrepresentation of proteins involved in signaling, such as DNA- and RNA-binding proteins as well as protein kinases (de la Fuente van Bentem et al., 2006; Sugiyama et al., 2008). This was expected and highlights the importance of regulatory phosphorylations in the control of cellular signaling. Although typical mammalian-type Tyr kinases have not been found in plants, Sugiyama et al. (2008) reported 4% of all identified phosphorylation events on Tyr. Interestingly, this level of Tyr phosphorylation is close to that

[1] This work was supported by a Marie Curie EST training network ADONIS (grant no. MEST–CT–2005–020232 to W.G. and S.B.) and by the 6th European Integrated Projects AGRON-OMICS (grant no. LSHG–CT–2006–037704 to W.G.). A.E. was supported by a Graduate Research Fellowship of the Zurich-Basel Plant Science Center to S.B.

* Corresponding author; e-mail sbaginsky@ethz.ch.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) are: Wilhelm Gruissem (wgruissem@ethz.ch) and Sacha Baginsky (sbaginsky@ethz.ch).

[W] The online version of this article contains Web-only data.
www.plantphysiol.org/cgi/doi/10.1104/pp.109.138677

expected for human cells (between 1.8% and 6%), suggesting that Tyr signaling may be similarly important in plants.

In vivo phosphorylation sites provide important information about the activity of protein kinases in their cellular context. Specific phosphorylation motifs reflect preferences of protein kinases for a certain amino acid context of the phosphorylation site. This has allowed the development of software tools for phosphorylation site prediction (for review, see Blom et al., 2004). Using phosphoproteomics, it is now possible to infer in vivo kinase activities from phosphorylation motifs. Motif-X (motif extractor) extracts significantly enriched kinase motifs from large-scale phosphoproteomics data (Schwartz and Gygi, 2005), thereby providing information about kinase/substrate relationships. Linding et al. (2007) systematically used phosphorylation motif occurrence for the construction of in vivo phosphorylation networks in human cells. A prediction of kinase/substrate relationships based exclusively on phosphorylation motifs resulted in accuracies of 19% to 34% for a benchmark kinase network. However, when context information such as protein-protein interaction and colocalization of kinase and a putative substrate was added to the prediction, the accuracy increased to 47% to 81%. Thus, phosphoproteome information is useful for the assembly of in vivo phosphorylation networks, provided that context information is available to constrain the number of possible kinase/substrate associations.

We performed a large-scale phosphoproteome analysis of Arabidopsis shoots, from which we assembled the chloroplast phosphoproteome. To provide insights into the dynamics of chloroplast protein phosphorylation during the circadian cycle, we also analyzed shoot phosphoproteomes at the end of day and end of night. Although much is known about the role of phosphorylation in the regulation of photosynthesis and chloroplast gene expression (for review, see, Link, 2003; Bollenbach et al., 2004; Rochaix, 2007), no comprehensive information is currently available for the chloroplast phosphoproteome. A small number of chloroplast protein kinases and their substrates have been identified. For example, the thylakoid-associated kinases STN7 and STN8 phosphorylate light-harvesting complex subunits as well as photosystem core subunits (for review, see Rochaix, 2007; Eberhard et al., 2008). Additionally, a casein kinase II (CKII)-like enzyme has been reported to phosphorylate proteins involved in transcription and posttranscriptional regulation (for review, see Link, 2003; Bollenbach et al., 2004) and the ATP synthase β-subunit (ATPB) of chloroplast ATP synthase (Kanekatsu et al., 1998).

The comprehensive phosphoproteome analysis reported here provides new insights into chloroplast phosphorylation networks and phosphoproteins required for different chloroplast functions. Phosphorylation motif analysis allowed us to establish kinase/substrate relationships and to obtain insights into the extent of different chloroplast kinase activities. Our data suggest considerable cross talk between kinases and integration of different chloroplast functions by individual kinases or kinase families.

RESULTS AND DISCUSSION

The Arabidopsis Phosphoproteome

We analyzed the phosphoproteome of Arabidopsis shoots and rosette leaves using IMAC and TiO_2 phosphopeptide enrichment strategies in combination with high-accuracy mass spectrometric phosphopeptide detection in an LTQ-Orbitrap mass spectrometer (Fig. 1A). The resulting data files were analyzed using the MASCOT and INSPECT search algorithms (Tanner et al., 2005; Payne et al., 2008; www.matrixscience.com). We identified 1,429 phosphoproteins and 3,029 unique phosphopeptides at a false discovery rate of 0.44% for MASCOT and 0.12% for INSPECT searches and established the exact sites of phosphorylation in 2,349 peptides (Table I; Supplemental Tables S1–S3). Details of peptide assignment, together with the original spectra, are publicly available in the PRIDE database (www.ebi.ac.uk/pride/prideMart.do; Jones et al., 2008). Our data set is comparable in size to the phosphoproteome data set reported recently by Sugiyama et al. (2008). We found 88% of all phosphorylation events on Ser and 11% on Thr (Fig. 1B), which is comparable to the distribution reported by Sugiyama et al. (2008). In contrast, we observed Tyr phosphorylation in only 0.3% of the peptides, which is significantly lower than previously reported (Sugiyama et al., 2008). The discrepancy in the rate of Tyr phosphorylation can be explained by the fact that we used differentiated photosynthetic plant tissues for our analysis, while Sugiyama et al. (2008) used cultured Arabidopsis cells. It is conceivable that the rate of Tyr phosphorylation in cultured cells does not reflect the situation found in differentiated tissues. Furthermore, some of the putative Tyr phosphorylation sites reported by Sugiyama et al. (2008) may need to be reconsidered (de la Fuente van Bentem and Hirt, 2009).

We compared the phosphoprotein detection in the two data sets and found an overlap of 600 proteins that were identified in both studies, whereas 743 were exclusively detected by Sugiyama et al. (2008) and 829 were exclusively detected in our phosphoproteome analysis. Information about the identified proteins in the two different data sets is provided in Supplemental Table S4. In order to characterize the differences in protein phosphorylation in the different biological materials, we performed a topGO analysis to identify overrepresented functional classes of proteins. Assignment of the phosphopeptides in our data set to functional GO categories using topGO showed a strong overrepresentation of proteins in the cellular component categories Chloroplast thylakoid membrane, Plastoglobule, and Light-harvesting complex (Fig. 1B). Interestingly, the data set reported by Sugiyama

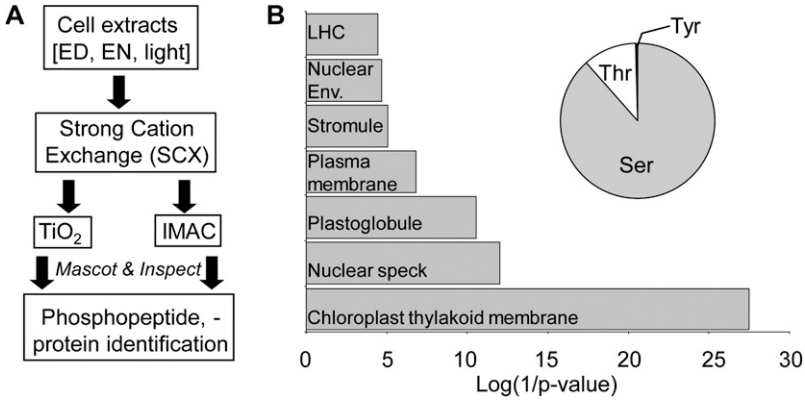

Figure 1. Acquisition and characteristics of the phosphoprotein data set. A, Workflow for the acquisition of phosphopeptides from different Arabidopsis plant samples. ED, End of day; EN, end of night, IMAC, immobilized metal affinity chromatography. B, Overrepresented GO categories in the complete phosphoproteome data set compared with the Arabidopsis TAIR8 protein database as determined by the elim method implemented in the topGO algorithm (Alexa et al., 2006). Shown are overrepresented categories with P values below 10^{-5} from the GO category Cellular component. The pie graph shows relative amounts of Ser, Thr, and Tyr phosphorylation in phosphopeptides for which the exact site of phosphorylation was determined.

et al. (2008) was enriched in nuclear and plasma membrane proteins. Also, topGO analysis of our data set revealed a significant enrichment of proteins involved in the molecular functions RNA metabolism, Intracellular trafficking, and Protein phosphorylation (Supplemental Table S5), supporting previous data that phosphopeptide enrichment is effective to detect low-abundance proteins involved in signaling (for review, see Preisinger et al., 2008).

Defining the Chloroplast Phosphoproteome

For the characterization of the plastid phosphoproteome, we decided to build a chloroplast proteome reference table instead of using targeted chloroplast proteomics for two reasons. First, it is well established that high phosphoproteome coverage depends on a significant amount of starting material, which is easier to obtain from intact cells (for review, see Kersten et al., 2006). Second, phosphoproteins are unstable during tissue extraction and protein isolation (Espina et al., 2008); therefore, rapid fractionation protocols increase phosphoprotein coverage and reproducibility. This is particularly important for a thorough analysis of phosphoproteome dynamics. Using a combination of protein-targeting prediction, organelle proteomics data, and complementary protein-targeting information available in the SUBA database (http://www.plantenergy.uwa.edu.au/suba2/; Heazlewood et al., 2007), we assembled a list of 1,619 high-confidence chloroplast proteins (Supplemental Table S6), of which 894 were identified in the flow-through fraction of the affinity chromatography and 174 were found to be phosphorylated in our study (Table I; Supplemental Tables S1 and S7). The assignment as chloroplast protein was based on at least one of four criteria. First, the proteins were detected in at least two different chloroplast proteomics studies; second, they were targeted to plastids as a GFP fusion protein; third, they were detected in at least one plastid proteome analysis and predicted to be localized to plastids by at least one prediction software tool; fourth, they were predicted to localize to plastids by at least three different prediction software tools. Based on these criteria, we identified 143 chloroplast phosphoproteins that were not present in the data set of Sugiyama et al. (2008). On the other hand, we did not detect 44 proteins that were reported by Sugiyama et al. (2008). This difference and the small overlap of 31 proteins that were found phosphorylated in both studies show that protein phosphorylation in plastids from cultured cells is considerably different from that of chloroplasts in photosynthetic tissues.

Altogether, we identified 353 phosphopeptides in 174 putative chloroplast proteins (Table I; Supplemental Table S1). Seventy-two percent of all detected phosphorylation events occurred at Ser, and 27% occurred at Thr. Tyrosine phosphorylation was suggested in the case of four peptides (1%), but the two search algorithms delivered contradictory results (Supplemental Table S2). INSPECT supports Tyr phosphorylation but MASCOT does not, suggesting that the Tyr phosphorylation site is questionable. INSPECT was trained on LTQ data and searches the tandem mass spectrometry (MS/MS) spectrum for sequence tags and the characteristic neutral loss of phosphoric acid (−98 atomic mass units [amu]) from the parent ion (Payne et al., 2008). Because phosphotyrosine is more stable than phosphoserine or phosphothreonine, the neutral loss does not readily occur, therefore causing problems for the assignment of Tyr phosphorylation. In order to characterize the phosphorylation of the four peptides in more detail, we subjected all spectra to de novo sequencing using PepNovo (Frank et al., 2007). None of the de novo sequencing results supports Tyr phosphorylation,

Table I. *Identified high-confidence chloroplast phosphoproteins*

Phosphorylated residues are marked with a p (e.g. pS or pT) in cases where the phosphorylation site could be assigned unambiguously. Oxidized Met is labeled as °M. In some cases, the exact site of phosphorylation could not be determined. These peptides are lacking any amino acid labeling (column Modification). Not included are those peptides and proteins that were exclusively identified by INSPECT but not by MASCOT. The full set of chloroplast phosphopeptides is provided in Supplemental Table S1.

Locus	Protein Description	Modification
Photosynthesis and Calvin cycle		
AT1G06680	PSBP-1 (oxygen-evolving enhancer protein 2)	FEDNFDATSNLNV°MVTPTDKK
AT1G15820	LHCB6 (light-harvesting complex PSII)	pTLIVAAAAAQPK
AT1G29910, AT1G29920	CAB3 (chlorophyll *a/b*-binding protein 3)	AVNLSPAASEVLGpSGR
		LSPAASEVLGSGR
		LSPAASEVLGpSGR
AT1G31330	PSAF (PSI subunit F)	LESSLKLYAPESAPALALNAQIEK
		KLESSLK
AT1G67090	RBCS1A (Ribulose-bisP carboxylase)	EHGNSPGYYDGR
		QVQCISFIAYKPPpSFTG
		EHGNpSPGYYDGR
AT1G79040	PSBR (PSII subunit R)	IKTDKPFGINGpS°MDLR
AT2G34430	LHB1B1 (PSII light-harvesting complex gene 1.4)	LSPAASEVFGpTGR
		KApSKPTGPSGSPWYGSDR
AT2G39730	RCA (Rubisco activase)	GLAYDTSDDQQDITR
		GLAYDpTSDDQQDITR
AT2G40100	LHCB4.3 (light-harvesting complex PSII)	FGFpSFGK
AT2G47910	CRR6 (chlororespiratory reduction 6)	LDLpSPFQR
AT3G04790	Rib-5-P isomerase-related	pSLGIPLVGLDTHPR
AT3G08940	LHCB4.2 (light-harvesting complex PSII)	NLYGEVIGTRTEAVDPK
		FGFGpTKK
		NLYGEVIGTRpTEAVDPK
		NLYGEVIGpTRTEAVDPK
AT3G12780	PGK1 (phosphoglycerate kinase 1)	SVGDLTSADLK
		pSVGDLTSADLK
AT3G47470	LHCA4 (PSI light-harvesting complex gene 4)	DLSFTSIGSSAK
		DLpSFTSIGSSAK
AT4G05180	PSBQ/PSBQ-2/PSII-Q (PSII subunit Q-2)	FYIQPLSPTEAAAR
		FYIQPLpSPTEAAAR
AT4G10340	LHCB5 (light-harvesting complex of PSII 5)	SKAVSETSDELAK
AT4G12800	PSAL (PSI subunit L)	SSFSSASLSQR
AT4G22890	PGR5-like A	ATTEQSGPVGGDNVDSNVLPYCSINK
AT4G28750	PSAE-1 (PSA E1 knockout; catalytic)	AAEDPAPASSSSKDSPAAAAAPDGATATKPKPPPIGPK
AT5G01530	Chlorophyll *a/b*-binding protein CP29 (LHCB4)	NLAGDVIGTRTEAADAK
		NLAGDVIGpTRpTEAADAK
		NLAGDVIGpTR
		NLAGDVIGTRpTEAADAK
AT5G64040	PSAN (PSI reaction center subunit PSI-N)	AFTVQFGSCKFPENFTGCQDLAK
ATCG00270	PSII D2 protein	pTIALGKFTK
ATCG00280	CP43 subunit of PSII	SPTGEVIFGGET°MR
ATCG00710	Encodes an 8-kD phosphoprotein (PSBH)	ATQTVEDSSRSGPR
		STTVGKLLKPLNSEYGK
		ATQTVEDSSR
		ATQpTVEDSSR
		ApTQpTVEDSSR
		ApTQTVEDSSR
AT2G28000	60-kD chaperonin α-subunit	NVVLDEFGpSPK
AT2G46820	PSAP/PSI-P/PTAC8/TMP14	ATTEVGEAPATTTEAETTELPEIVK
Gene expression and nucleic acid binding		
AT1G07320	RPL4 (ribosomal protein L4)	YGVDAVEEEDDDEDETEGpSEEA
AT1G30480	DRT111 (DNA damage-repair/toleration protein 111)	SSSPPGNVDGFSIGK
AT1G48620	HON5 (high mobility group family A5)	KDGTSPTVKPAASVSGGVETVK
		RVDAGGASSVAPPPPPPTNVESGGEEVAVK
AT1G79850	RPS17 (ribosomal protein S17)	TKSFVALPVIAR
AT2G35410	33-kD ribonucleoprotein	ETSADEETSQEEK
		SASESEDGDSVEANNASEDGDTVEDK

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Locus	Protein Description	Modification
AT2G37220	29-kD ribonucleoprotein	SSFGSSGSGYGGGGGSGAGSGNR
AT2G38140	PSRP4 (plastid-specific ribosomal protein 4)	IKIDIDESLFpSN
AT3G03710	3′-5′-Exoribonuclease/RNA binding	ALLPESETDKDSQK
		ALLPEpSETDKDSQK
AT3G46780	PTAC16 (plastid transcriptionally active 18)	LGSQFATAIQNASETPK
		DISSGLSWNKLGSQFATAIQNASETPKVQVATVR
		LGSQFATAIQNASETPKVQVATVR
		EAEAApSLAEDAQQK
		LGSQFATAIQNASETPKVQVApTVR
		VQVApTVR
AT3G48500	PDE312/PTAC10 (pigment defective 312)	LSELSDDEDFDEQK
		KLSELpSDDEDFDEQK
AT3G53460	CP29 (chloroplast 29-kD ribonucleoprotein)	REEpSFSR
AT3G63140	mRNA-binding protein, putative	STEQPPHVEGDAVK
AT4G13670	PTAC5 (plastid transcriptionally active 5)	LF°MDEDVEpTDKDEAST°MKK
AT5G08610	DEAD box RNA helicase (RH26)	GKFTSDEDNADPEVVR
		FTSDEDNADPEVVR
		GKFpTpSDEDNADPEVVR
AT5G24490	30S ribosomal protein, putative	VREPVIEPVVEDVEDSTDSSVGEEEEEDDLIK
AT5G65220	Ribosomal protein L29 family protein	LQEEEAAEEAAEAAKpSA
ATCG01240, ATCG00900	30S chloroplast ribosomal protein S7	VGGSTHQVPIEIGSTQGK
Transporter		
AT1G01790	KEA1 (K efflux antiporter 1)	IGESSESSDETEATDLK
		SLSISQTPEETQGQLSDEETSQEDA°MVLSGNVEDVTHQVEK
AT1G06950	ATTIC110/TIC110	LANAVSSGDLEAQDSK
AT1G15500	ATP:ADP antiporter	ASSVKIPVVSQEDAPSGETTSQLSEK
AT1G59870	PDR8/PEN3 (pleiotropic drug resistance 8)	LRTTL°MNAVVEDDVYGNQL°MSK
		NIEDIFSSGSR
		SLSTADGNR
		NIEDIFSSGpSRR
		SLpSTADGNR
AT1G80300	ATP:ADP antiporter	ASSVKIPVVSQDESGNGSLGESPSSSPEK
AT2G32040	Integral membrane transporter family protein	ISVAEGDTSNTDVEGDRDTTSSIR
		RRDpSEESLLLDSR
AT3G48890	ATMP2 (membrane steroid-binding protein 2)	TASAEGLSTNTGEEASAITHDETSR
		DVApTDDDDAAKE
AT4G00630	KEA2 (K^+ efflux antiporter)	STSKPKPPSPSETSDDNQIIEGTLAI
AT4G02510	TOC159	FTSESDSIADSSK
		ASSGIEAHSDEANISNNMSDR
		EVDQEPSGEGVTRVDGSESEEETEE°MIFGSSEAAK
		IDGQIVTDSDEDVDTEDEGEEK
		INADAETLEVANKFDQIGDDDSGEFEPVSDK
		ELDSSSEAVSGNSDKVGADDLSDSEK
		VGADDLSDSEK
		VDGSESEEETEE°MIFGSSEAAK
		FDQIGDDDSGEFEPVpSDK
		FDPIGQGEGGEVELEpSDK
		IDGQIVpTDpSDEDVDpTEDEGEEK
		KVVEGDpSAEEDENK
		FDQIGDDDpSGEFEPVSDK
		IDGQIVpTDpSDEDVDpTEDEGEEK
		VDVDDKpSDNVIEEEGVELTDK
		VGADDLpSDSEK
		EFpSFGGKEVDQEPSGEGVTR
		VDGpSEpSEEETEE°MIFGSSEAAK
AT5G13550	Sulfate transporter 4.1	YGGSNNNSSSSNALLKEPLLSVEK
AT4G26670	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	AANDSSNAIDIDGNLDSDSNLNTDGDEATDNDSSK

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Locus	Protein Description	Modification
Metabolism and energy		
AT1G17745	PGDH (3-phosphoglycerate dehydrogenase)	FSTVGSDSDEYNPTLPKPR
AT1G78510	SPS1 (solanesyl diphosphate synthase 1)	QFPpSLAK
AT2G05710	Aconitate hydratase	TFSS°MASEHPFK
AT2G38040	CAC3 (acetyl-CoA carboxylase carboxyltransferase α-subunit)	ELAAEESDGSVK
		ELAAEESDGSVKEDDDDDEDSSESGK
		ELAAEEpSDGSVK
AT2G39800	P5CS1 (Δ1-pyrroline-5-carboxylate synthase 1)	QLVNSSFADLQKPQTELDGK
AT3G01180	ATSS2 (starch synthase 2)	VEASGSDDDEPEDALQATIDK
AT3G01500	CA1 (carbonic anhydrase 1)	VEQITAALQTGTSSDKK
AT3G18680	Asp/Glu/uridylate kinase family protein	LPSFDGTSKPPLK
AT3G57610	ATPURA (adenylosuccinate synthase)	LAGQEFGTTTGRPR
AT3G60750	Transketolase, putative	ALPTYTPESPGDATR
		ALPTYTPEpSPGDATR
AT4G04640	ATPC1 (ATP synthase γ-chain 1)	SLS°MVYNR
		SLpS°MVYNR
AT4G10750	HpcH/HpaI aldolase family protein	°MGLVNESDSEDSSEHDK
AT4G14070	AAE15 (acyl-activating enzyme 15)	EKEVKPSSPFLESSSFSGDAALR
		EVKPSSPFLESSSFSGDAALR
AT4G15530	PPDK (pyruvate orthophosphate dikinase)	GG°MTSHAAVVAR
		GG°MTpSHAAVVAR
		GG°MpTSHAAVVAR
AT4G18480	CHLI1 magnesium chelatase	VCpSELNVDGLR
AT4G24620	PGI1 (chloroplastic phosphoglucose isomerase)	VLIAEGNCGpSPR
AT4G32260	ATP synthase family	ALDpSQIAALSEDIVKK
AT5G01220	SQD2 (sulfoquinovosyldiacylglycerol 2)	EDDEpSEIDAPLLDPESLSKPR
AT5G08280	HEMC (hydroxymethylbilane synthase)	ILpSQPLADIGGK
AT5G11670	ATNADP-ME2 (NADP-malic enzyme 2)	GSTPTDLPGEDVADNR
AT5G14200	3-Isopropylmalate dehydrogenase	AGpSLEGLEFDFK
AT5G14740	CA2 (β-carbonic anhydrase 2)	GNESYEDAIEALK
		ITAELQAASSSDSK
		VLAESESSAFEDQCGR
		GNEpSYEDAIEALKK
		VLAESEpSSAFEDQCGR
AT5G19220	ADG2 (ADPG pyrophosphorylase 2)	VGpSNVQLK
AT5G36790, AT5G36700	Phosphoglycolate phosphatase, putative	ISDFLpSPK
AT5G36880	Acetyl-CoA synthetase	HVES°MSQLPSGAGK
AT5G51820	PGM (phosphoglucomutase)	ANGGFIMSASHNPGGPEYDWGIK
AT5G64300	ATGCH (*A. thaliana* GTP cyclohydrolase II)	FKGDVVEKIESESES
ATCG00120	Encodes the ATPase α-subunit	GKISApSESR
AT5G63310	NDPK2 (nucleoside diphosphate kinase 2; ATP-binding/nucleoside diphosphate kinase)	NIVHGSDpSPENGKR
ATCG00480	Chloroplast-encoded gene for β-subunit of ATP synthase	TNPTTpSNPEVSIR
		TNPTTSNPEVpSIR
		TNPTTSNPEVSIR
Other or unknown function		
AT1G08640	Heat shock protein binding	GVTFGSFK
		GVTFGSFKVSK
		GVTFGpSFKVSK
AT1G19870	IQD32 (IQ-domain 32)	ETLESALLKSPSPDNNNVSEK
		IEEDVTSEVE°MASK
		ITSSPKQEIGTGEATEQEEGKEQK
		RTSFGYDQEAR
		SASQAQQGTKDR
		TRETLESALLKSPSPDNNNVSEK
		KVSNPSFIAAQSK
		VEPEESESDDVIIVR
		TRETLESALLKpSPpSPDNNNVSEK

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Locus	Protein Description	Modification
		TRETLESALLKSPpSPDNNNVSEK
		pSDAEGAEPR
		TpSFGYDQEAR
		VEPEEpSESDDVIIVR
		KVpSNPSFIAAQSK
		HSLPGVTNGKQVpSPR
AT1G22700	Tetratricopeptide repeat (TPR)-containing protein	LYKGVVAVKpSK
AT1G33810	Similar to unknown protein (GB:ABK94119.1)	APWRGDDEDDpSDKFSNAK
AT1G54520	Similar to unknown (GB:ABK96363.1)	IGGNSFSSR
		SSSSSSSQSYSVPR
		IGGNpSFSSR
AT1G55490	CPN60B (chaperonin 60β)	LApSKVDAIKATLDNDEEK
AT1G58200	MSL3 (MSCS-like 3)	DEVpSDDEATIEQTLK
AT1G72640	Binding/catalytic	DGSLSGDDDEFEEEK
AT1G74730	Similar to unknown protein (GB:ABK96654.1)	AEGLGLTLSSLEK
AT1G76080	ATCDSP32/CDSP32 (chloroplastic drought-induced stress protein of 32 kD)	S°MpSETVVFAR
AT1G76180	ERD14 (early response to dehydration 14)	SDSSSSSSSEEEGSDGEKR
		VHIpSEPEPEVKHESLLEK
AT2G01870	Similar to unnamed protein product	IINKTSSDVVR
AT2G28800	ALB3 (albino 3)	DTVELVEESQSESEEGSDDEEEEAR
		AVAKDTVELVEESQSESEEGSDDEEEEAR
AT2G30930	Similar to unknown protein (TAIR:AT1G06540.1)	LVGGVTNLVSGASSSTVANR
		ATSALSEAK
		ATpSALSEAK
		pSLLQTFEAK
AT2G37080	Myosin heavy chain-related	TGSLEpSPLR
AT2G37660	Unknown protein	ALDLASKPEGTGpTPTK
AT2G47400	CP12-1	ATSEGEISEK
		ATpSEGEISEKVEK
AT3G09050	Similar to unknown protein (GB:ABK96465.1)	SGDGTSDSDSDPDPPKPEGDTRR
AT3G13470	Chaperonin, putative	LApSKVDAIKDTLENDEEK
AT3G16000	MFP1 (MAR-binding filament-like protein 1)	EARKpSLETDLEEAVK
AT3G18390	EMB1865 (embryo-defective 1865)	NLGLGpSDDEDDVEDDEGGGINGGDVKPVTGEER
AT3G18890	Binding/catalytic/coenzyme binding	KSDSLSPGPTDSDTDKSSTVAK
AT3G20550	DDL (dawdle)	AIASRHDEGSNARGGSEEPNVEEDSVAR
		HDEGSNARGGSEEPNVEEDSVAR
		GGpSEEPNVEEDSVAR
AT3G22520	Similar to unknown protein (TAIR:AT4G14840.1)	DSSVLVSVSpSP°MR
AT3G23400	Plastid-lipid-associated protein PAP	GLVApSVDDLER
AT3G25690	CHUP1 (chloroplast unusual positioning 1)	SFYGGSPGRLSSS°MNK
		TQQQASSPGEGLNSVAASFHV°MSK
		SKDDSSVQSSPSR
		DLSKNLpSPK
		L°MLEYAGpSER
		SFYGGpSPGR
AT3G26740	CCL (CCR-LIKE)	EYEEYNpSPK
AT3G47070	Similar to unknown (GB:ABK95428.1)	VDEKEGTTTGGR
		VDEKEGTTTGGRGTVR
		EGTTTGGRGTVR
		KSEGGFGGLGpSLFKK
		GGpSGGPKEEKNPIDFVLGF°MTK
		EGpTTTGGRGpTVR
		EGTTTGGRGpTVR
		EGpTTTGGRGTVR
AT3G52230	Similar to unknown protein (GB:ABK93315.1)	EATGDDDQKDDDEDDQSSDGHED
AT3G53130	LUT1 (lutein-deficient 1)	TLSSGKNDESGIPIANAK
AT3G59400	GUN4 (genomes uncoupled 4)	VFKTNYpSF
AT3G63160	Unknown protein	LSSSKDKSDSDDATVPPPSGA
		DKSDSDDATVPPPSGA
		DKSDSDDApTVPPPSGA

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Locus	Protein Description	Modification
		DKSDpSDDATVPPPSGA
		VEKpSGGEVNFPK
AT4G01150	Uncharacterized protein	ASSEETSSIDTNELITDLK
AT4G04020	FIB (fibrillin)	ApTDIDDEWGQDGVER
AT4G09160	SEC14	SLGSFKEETNK
		SLGpSFKEETNK
AT4G14870	P-P bond-hydrolysis-driven protein transmembrane transporter	DTAGSESESEATPSPAEESGSGEDKEVEISAIGAEIK
AT4G15560	CLA1 (cloroplastos alterados 1)	LQpSNPALR
AT4G17620	Gly-rich protein	DLFGpSDNEEYTK
AT4G19100	Similar to unknown protein (TAIR:AT5G52780.1)	SKKPKPGNQSDEDDDDEDEDDDDEEDER
AT4G23890	Similar to unnamed protein product (GB:CAO69542.1)	ESDDSNLpSFKVPEDGFEK
AT4G35770	SEN1 (dark-inducible 1)	NPpSFLR
AT4G36970	Remorin family protein	AE°MTTA°MQpSPVVSRR
AT4G39040	Similar to unknown protein (TAIR:AT2G21350.1)	GIEDSEASEEVSEIGDKEEK
AT5G08050	Similar to unnamed protein (GB:CAO62462.1)	SGFSLSTIERLGLLTK
		SGFSLSTIER
AT5G08540	Similar to unnamed protein (GB:CAO65975.1)	KNSSVEEETEEEVEED°MPWIQEK
		TAESSSDKEEDSNEEDDSNTTS
AT5G17170	ENH1 (enhancer of SOS3-1)	KLpTETQKAR
AT5G22640	EMB1211 (embryo-defective 1211)	TE°MGLTEEDEDVLVPVYKEEK
		KTE°MGLpTEEDEDVLVPVYKEEK
		KpTE°MGLTEEDEDVLVPVYK
		DGEQpSPGGSLTPPQK
AT5G23060	Similar to unknown protein (TAIR:AT3G59780.1)	IIPAASRSFGTR
		LGTDSYNFSFAQVLSPSR
		LGTDSYNFSFAQVLSPSRIIPAASR
		SFGTRSGTKFLPSSD
		SGTKFLPSSD
		pSFGTRSGpTKFLPSSD
		pSGTKFLPSSD
		ELESSKSPVPESTDGSKDELNIYSQDELDDNR
AT5G26742	EMB1138 (embryo-defective 1138)	SLGLSDHDEYDLDGDNNNVEADDGEELAISK
		pSFGGSCFICGK
AT5G63420	EMB2746 (embryo-defective 2746)	DDDELADASDSETKSSPK
		ENSRDDDELADASDSETKSSPK
		DDDELADASDSETK
		ENSRDDDELADASDSETK
		ENSRDDDELADApSDSETK
ATCG01130	Hypothetical protein	KLSFFSEPQQEEK
Kinases and proteases		
AT1G66670	CLPP3 (Clp protease proteolytic subunit 3)	V°MIHQPLGpTAGGK
AT1G68830	STN7 (STT7 homolog STN7)	TVTETIDEISDGRK
		TVpTETIDEISDGRK
		LVKpTVTEpTIDEISDGRK
		LVKpTVTETIDEISDGRK
		NALApSALR
AT5G40200	DEGP9 (DEGP protease 9)	EASANEASLPQSPEPVSASEANPSPSRR
AT5G42270	VAR1 (variegated 1)	SKFQEVPETGVTFGDVAGADQAK
		SKSKFQEVPETGVTFGDVAGADQAK
AT5G53170	FTSH11 (FtsH protease 11)	LVSDTEVSELETNDRFVGGEETK

and the MS/MS spectrum quality was not sufficiently high to support an unambiguous sequence assignment (Supplemental Spectra S1). This finding was supported by manual data interpretation. Thus, our phosphoproteome data set does not provide unambiguous evidence for Tyr phosphorylation in chloroplasts. Further experiments are necessary to validate or disprove the existence of phosphotyrosine in plastids.

Functionality of Chloroplast Protein Phosphorylation

Chloroplasts are photosynthetically active and therefore a source of reactive oxygen generated by

water oxidation. To avoid damage caused by reactive oxygen, plants have evolved posttranslational signal transduction and control mechanisms that facilitate the rapid adaptation of chloroplast functions to prevailing light conditions. Therefore, we were interested to understand the regulatory chloroplast phosphorylation network in more detail. We found phosphorylated chloroplast proteins that have different functions (Table I). The largest groups of phosphorylated proteins are involved in chloroplast gene expression and photosynthesis.

It has been proposed that phosphorylation of chloroplast RNA polymerase decreases the rate of transcription elongation (Tiller and Link, 1993; Link, 2003). We found that the transcriptionally active chromosome (TAC) subunits TAC16, TAC10, and TAC5 were phosphorylated. The phosphorylation of RNA-binding proteins has been implicated in the stabilization of plastid mRNAs (Baginsky and Gruissem, 2002; Loza-Tavera et al., 2006). We identified RNP29 and RNP33 as chloroplast phosphoproteins. According to a mechanistic model of plastid mRNA stability (for review, see Bollenbach et al., 2004), phosphorylation of the two proteins increases their affinity for RNA. We suggest that this results in stable RNP/RNA complexes, which protect the mRNA from endonucleolytic cleavages and allow translation of more protein from an individual mRNA molecule. Our view is consistent with the light-dependent rapid phosphorylation of RNP29 and RNP33 during the early phase of deetiolation, when the massive reorganization of the plastid proteome to support photosynthesis requires higher translation efficiencies (Kleffmann et al., 2007).

Proteins involved in photosynthesis constitute another large group of phosphorylated proteins. We confirmed the known phosphorylation of abundant light-harvesting and photosystem subunit proteins. Light-harvesting complex (LHC) and photosystem core subunit phosphorylation occurs predominantly, but not exclusively, on Thr. We detected Ser phosphorylation in chlorophyll-binding proteins 2/3 and LHC subunits (Table I). In addition, we found the thylakoid-associated kinase STN7 to be an abundant thylakoid phosphoprotein, which has not been reported previously. STN7 catalyzes the redox-dependent phosphorylation of LHCII subunits to regulate photosynthetic state transitions (Bellafiore et al., 2005). We identified four phosphorylation sites in the C-terminal region of STN7. Interestingly, the phosphorylated C terminus of the protein is not conserved between *Chlamydomonas* and Arabidopsis (Rochaix, 2007). This suggests that the C-terminal phosphorylation of STN7 controls responses that are specific for higher plants (e.g. long-term adaptation to environmental conditions that are not relevant for *Chlamydomonas*). In addition to regulatory phosphorylations of photosynthetic proteins, we identified nine phosphorylated proteins that function in chloroplast energy and carbohydrate metabolism, including phosphorylation of the chloroplast ATPB (Table I). ADP-Glc pyrophosphorylase is a highly regulated enzyme that controls starch accumulation in photosynthetic tissues (for review, see Geigenberger et al., 2005). The single phosphorylation site we detected in ADP-Glc pyrophosphorylase suggests that in addition to other regulatory mechanisms, starch synthesis is controlled by the phosphorylation status of the enzyme.

Chloroplast Phosphoproteome Dynamics

The interaction between starch, carbohydrate, and energy metabolism is specifically affected by the activity of the photosynthetic electron transport and thus by the availability of light. In order to understand the role of phosphorylation in the control of these complex interactions, we next analyzed whether the phosphorylation status of chloroplast proteins changes during a light/dark cycle. To determine chloroplast phosphoproteome dynamics, we extracted phosphopeptides at end-of-day (1 h before the light was turned off) and end-of-night (1 h before the light was turned on) time points and analyzed the phosphoproteome in three biological replicates. Sequence alignment of the identified phosphorylation sites showed only minor differences and did not reveal a significant preference in phosphorylation site and phosphorylation motif utilization, as indicated by Motif-X analysis (data not shown). This suggests that end-of-day and end-of-night chloroplast protein phosphorylations are largely similar. Consistent with this conclusion, we found that phosphorylation of most chloroplast proteins was independent of light at these time points (Fig. 2; Supplemental Table S8).

LHC proteins LHCA4 and LHCB6 and chlorophyll *a/b* binding protein 3 were found phosphorylated

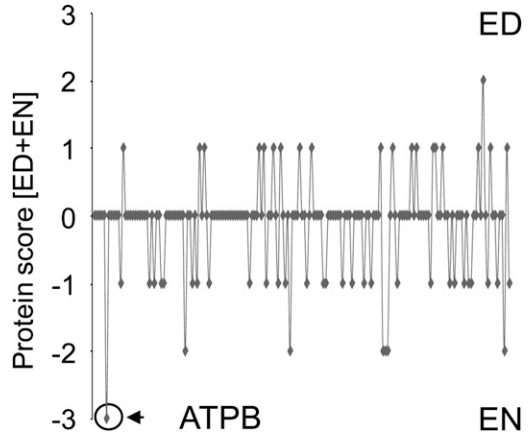

Figure 2. Identification of phosphoproteins at the end of day and end of night. Reproducibility score for end-of-day (ED)/end-of-night (EN) phosphorylations. Each end-of-day phosphorylation event was scored as +1, and each end-of-night phosphorylation event was scored as −1. All individual scores for three biological replicates were added to result in the final score plotted on the *y* axis. Only ATPB received the highest possible end-of-night score of −3, since it was found phosphorylated exclusively at the end of night in all biological replicates.

exclusively at the end of day, supporting their light-dependent phosphorylation by kinases associated with the thylakoid membranes (Supplemental Table S8). In contrast, phosphorylation of carbohydrate metabolism enzymes was not altered with the exception of starch synthase, which was found phosphorylated exclusively at the end of the dark period. Interestingly, ATPB was also found exclusively phosphorylated at the end of night (Fig. 2; Supplemental Table S8), confirming earlier suggestions that ATP synthase activity is controlled by phosphorylation (Kanekatsu et al., 1998; Bunney et al., 2001). The phosphorylation of ATPB is the only phosphorylation event that is reproducibly detected in three biological replicates only at the end of night and not at the end of day. Quantitative analyses suggested that ATPB phosphorylation occurs preferentially, but not exclusively, at the end of night (Fig. 3, A and B). The ATPB phosphopeptide precursor ion at 748.3438 amu is also detected in end-of-day samples, but at consistently lower intensity compared with the end-of-night samples. This is supported by the relative quantification results from SuperHirn, a software tool that quantifies peptides by their extracted ion chromatograms after aligning the liquid chromatography-mass spectrometry (LC-MS) runs (Mueller et al., 2007). The quantification results are presented in Supplemental Table S9. The comparison with nonplastid reference phosphopeptides that are not affected by the light/dark treatment supports the lower relative intensity of the ATPB precursor ion in end-of-day samples (Fig. 3B). Normalized spectral count quantification (Mueller et al., 2008) of the unphosphorylated ATPB protein in the flow-through fraction after affinity chromatography showed that the lower ATPB precursor intensity is a consequence of an altered ATPB phosphorylation state and is not due to fluctuations in ATPB concentration. The latter remains largely constant during a day/night cycle (Fig. 3C; Supplemental Table S10).

The Chloroplast Phosphoproteome Is Controlled by Different Protein Kinases

Confirmed phosphorylation sites are footprints of kinase activities. Although several chloroplast protein kinases have been identified, their substrate spectrum and functional interactions are mostly unknown (Schliebner et al., 2008). Therefore, we analyzed our phosphopeptide data set for putative phosphorylation motifs using Motif-X. Motif-X is a software tool that extracts overrepresented patterns from any sequence data set by comparison with a dynamic statistical background (Schwartz and Gygi, 2005). We expect that phosphorylation motifs reveal information about kinase activities and provide new insights into the chloroplast phosphorylation network. Contextual information for colocalization of kinase and substrate in the chloroplast that we have available from our data will further increase confidence in phosphorylation network assembly (Linding et al., 2007).

Known chloroplast protein kinases include CKII, two enzymes that resemble second messenger-dependent kinases, and thylakoid membrane-associated Ser/Thr kinases that cannot be classified into a single kinase family (Leister and Kleine, 2008; Schliebner et al., 2008). Recently, a chloroplast sensor kinase (CSK) was reported that has homology to bacterial two-component regulators (Puthiyaveetil et al., 2008). Except for CSK, all of the previously characterized chloroplast kinases are Ser/Thr kinases. CSK autophosphorylation is resistant to acid and alkali treatment, which was considered indirect evidence for Tyr phosphorylation. As described above, Tyr phosphorylation in plastids remains uncertain, and further

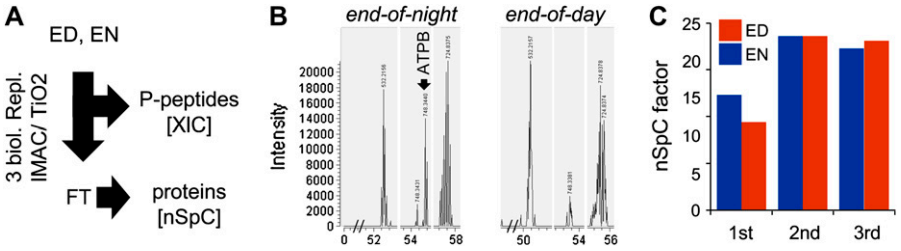

Figure 3. Quantification of ATPB phosphorylation at the end of day and end of night. A, Strategy for the quantification of the ATPB phosphopeptide and the ATPB protein. End-of-day (ED) and end-of-night (EN) samples are subjected to affinity chromatography on IMAC or TiO_2 as described in "Materials and Methods." B, Phosphopeptides are eluted from the affinity column and identified by mass spectrometry. The relative quantification of phosphopeptides in the different samples is based on their extracted ion chromatograms (XIC), and the ATPB phosphopeptide is presented in comparison with two reference peptides (532.215 amu, AT1G68060-MAP70 and 724.837 amu, AT4G15930-dynein light chain). Quantification was also based on SuperHirn (Mueller et al., 2007), which supported the consistently higher abundance of the ATPB phosphopeptide in all end-of-night samples (Supplemental Table S4). C, The unphosphorylated peptides are collected in the flow-through fraction and used for the quantification of proteins by their unphosphorylated peptides via normalized spectral counting (nSpC) in three biological samples.

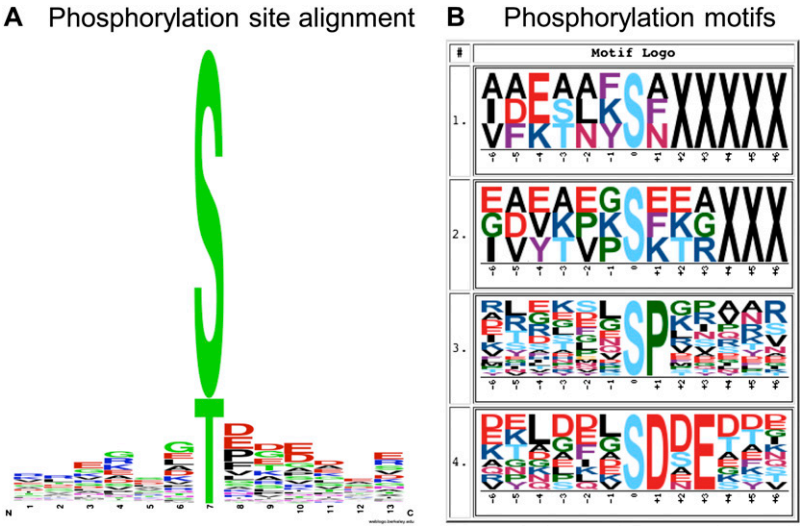

Figure 4. Sequence alignment of phosphorylation sites and extraction of significantly enriched phosphorylation motifs. A, Amino acid sequence around the phosphorylated amino acid based on alignment of all phosphorylation sites established in the entire chloroplast data set. B, Motif-X-extracted motifs from the entire phosphopeptide data set. The TAIR8 protein database was used as the background database to normalize the score against a random distribution of amino acids. Note that only those phosphorylated amino acids that were confidently identified as the exact site of phosphorylation were used for the analysis (see "Materials and Methods" for a description). Motifs 1 and 2, Unknown phosphorylation motifs (n = 3 + 3); motif 3, Pro-directed kinase motif (n = 21); motif 4, CKII motif (n = 10).

experiments are necessary to confirm CSK autophosphorylation at Tyr.

Motif-X extracted four significantly enriched phosphorylation motifs from all identified chloroplast phosphopeptides (Fig. 4). A striking enrichment of phosphorylation sites was detected in plastid ribosomal proteins proximal to their C termini (Fig. 4; Table I). The C-terminal motifs yielded the highest enrichment scores and might represent accessible Ser residues in ribosomal proteins assembled in the ribosome. Villen et al. (2007) previously observed phosphorylation at the C terminus of proteins in mouse and proposed that the occurrence of this motif may indicate kinase preference for acidic substrates, because the C terminus is in close proximity to the phosphorylation site. The other enriched motifs resemble known motifs from Pro-directed kinases and CKII. The Ser/Pro phosphorylation motif was originally described for substrates of mitogen-activated protein kinase and cyclin-dependent kinases (Adams, 2001), but other kinases phosphorylating this motif have been identified (for review, see Adams, 2001). We found several enzymes in plastid metabolism to be phosphorylated at this motif (e.g. phosphoglucose isomerase, nucleoside diphosphate kinase 2,2-isopropylmalate synthase, and transketolase), but the kinase that catalyzes phosphorylation at these sites remains elusive.

More functional information is available for CKII, which appears to be a major chloroplast kinase, because the alignment of all identified chloroplast phosphorylation sites is dominated by acidic residues characteristic for CKII substrate recognition (Fig. 4). CKII substrate preference for acidic amino acids around the Ser phosphorylation site was attributed to basic amino acids and His in the catalytic domain, which distinguish CKII from other kinases (Pinna, 1990). Biochemical characterization as well as mass spectrometric identification confirmed AT2G23070 as the only plastid CKII-like enzyme (Ogrzewalla et al., 2002; Salinas et al., 2006). The enzyme is involved in the regulation of chloroplast gene expression, and a number of its protein substrates involved in transcription and mRNA stability were reported previously (for review, see Link, 2003; Bollenbach et al., 2004; Lopez-Juez and Pyke, 2005). We identified several additional proteins involved in transcriptional and posttranscriptional regulation as potential CKII substrates, including TAC5, TAC10, and TAC16, as well as two RNA-binding proteins (Table I). Our analysis, therefore, significantly expands the current view of CKII-mediated phosphorylation control of plastid gene expression (Lisitsky and Schuster, 1995; Baginsky et al., 1997; Kanekatsu et al., 1998). CKII substrates are not restricted to gene expression, however, because metabolic enzymes such as carbonic anhydrase and ATP synthase are

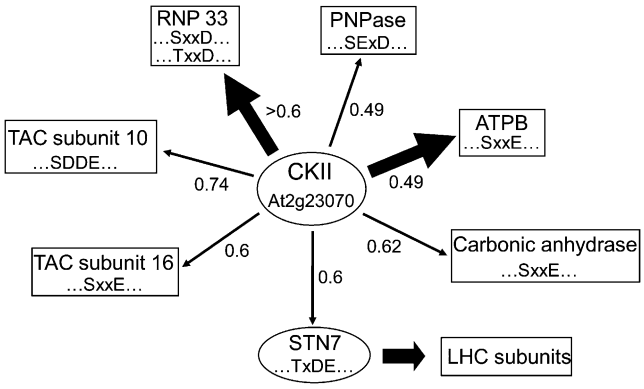

Figure 5. A phosphorylation network for chloroplast CKII. The network was assembled using NetPhosK motif analysis (Blom et al., 2004). The width of the arrows indicates the extent of additional evidence for CKII phosphorylation. Thick arrows are used in cases where phosphorylation has been biochemically characterized in vitro and inhibitor studies support CKII activity. Thin arrows indicate kinase/substrate relationships that are exclusively inferred from motif prediction. The motif score calculated by NetPhosK was assigned to the arrows. In all cases, the CKII motif received the highest score among all eukaryotic kinases that are considered by the NetPhosK algorithm (Blom et al., 2004).

also phosphorylated at CKII-type phosphorylation motifs (Fig. 5). This suggests that cross talk exists between chloroplast gene expression and metabolism, which is controlled by CKII. Furthermore, we identified one CKII phosphorylation site in STN7, suggesting that CKII may integrate regulatory phosphorylation cascades through phosphorylation of other kinases (Fig. 5).

Our data also provide evidence that ATPB is preferentially phosphorylated at CKII sites in vivo at the end of night (Table I; Fig. 2; Supplemental Tables S9 and S10). ATPB was established as a substrate for CKII in vitro (Kanekatsu et al., 1998), but it remained unclear whether such a phosphorylation could also occur in vivo. Although it was recently shown that ATPB is a phosphoprotein in vivo, the exact phosphorylation sites remained unknown (del Riego et al., 2006). Chloroplast ATP synthase switches its catalytic activity from ATP synthesis to ATP hydrolysis when the electrochemical gradient across the membrane collapses in the dark. In order to save ATP in the dark or during the night, ATP synthase becomes inactivated. It has been suggested that 14-3-3 proteins are involved in the inactivation of ATP synthase by phosphorylation-dependent interaction with the β-subunit (Bunney et al., 2001). In its phosphorylated form, ATPB can interact with 14-3-3 proteins, thereby preventing the rotation and catalytic action of the ATP synthase complex. Based on our findings and published results, we propose that CKII preferentially catalyzes ATPB phosphorylation in the dark. This scenario explains why ATP synthase activity is rapidly shut off in intact chloroplasts in the dark, whereas isolated thylakoid membranes retain ATP synthase activity for several hours, even in dark conditions (for review, see Richter et al., 2005).

CONCLUSION

Our analysis provides comprehensive information about the phosphoproteome of photosynthetically active Arabidopsis chloroplasts. A comparison with large-scale data obtained from heterotrophic Arabidopsis cell cultures revealed significant differences and suggests that a functional analysis of chloroplast protein phosphorylation must be performed with differentiated and functionally specialized plant tissues. Our data, therefore, provide new insights into the complexity of the chloroplast phosphoproteome and establish the basis for targeted phosphoproteomics, which will allow the full characterization of kinase/substrate relationships and phosphoproteome dynamics in response to diverse environmental signals. Further experiments are now required to unravel the substrate spectrum of the different chloroplast protein kinases and their phosphorylation network. A promising way forward is the comprehensive characterization of the chloroplast phosphoproteome from different kinase mutants with specialized targeted quantitative proteomics tools.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis (*Arabidopsis thaliana* ecotype Columbia-0) seedlings were grown under short-day conditions in a controlled-environment chamber (8 h of light/16 h of dark, 100 μE m^{-2} s^{-1}). Whole shoots were harvested after 22 d (light). For the comparative study of phosphorylation at the end of night and end of day, seedlings were grown under 12-h-light/12-h-dark conditions (100 μE m^{-2} s^{-1}) and harvested after 25 d. Rosette leaves were harvested 1 h before the light was turned off for the end-of-day sample and 1 h before the light was turned on for the end-of-night sample. The end-of-night samples were harvested under green light. All plant material was immediately frozen in liquid nitrogen and stored at −80°C until further analyses.

Protein Extraction

Protein extraction was done in two steps. First, soluble proteins were extracted from frozen and ground plant material by adding 40 mM Tris-HCl, pH 8, 5 mM MgCl, 1 mM dithiothreitol (DTT), and inhibitor cocktail for proteases (EDTA free; Roche) and phosphatases (phosSTOP; Roche). Second, 40 mM Tris, 4% SDS, and 40 mM DTT were added to extract membrane-associated and integral membrane proteins. Soluble proteins were precipitated by adding 5 volumes of ice-cold 80% acetone and 3 h of incubation at −20°C, whereas in the end-of-day and end-of-night samples, 10% TCA in acetone was used. Membrane proteins were precipitated by methanol/chloroform (Wessel and Flugge, 1984). Finally, both protein pellets were resolved separately in a small volume of resuspension buffer (20 mM Tris-HCl, pH 8.3, 3 mM EDTA, and 8 M urea). For tryptic digestion, the protein solution was diluted to 1 M urea by adding 20 mM Tris-HCl, pH 8.3, and 3 mM EDTA.

In-Solution Tryptic Protein Digest

Before tryptic digest, Cys residues were reduced with 10 mM DTT for 45 min at 50°C and alkylated with 50 mM iodoacetamide for 1 h at room temperature in the dark. Trypsin (Promega; sequencing grade) was added at a ratio of 1:20 and incubated overnight at 37°C.

Fractionation of Peptides by Strong Cation-Exchange Chromatography

Peptides were desalted using Sep-Pak reverse-phase cartridges (Waters), dissolved in buffer A (10 mM KH_2PO_4, pH 2.6, in 25% acetonitrile [ACN]), and loaded onto a 4.6- × 200-mm polySULFOETHYL aspartamide A column (PolyLC) on an Agilent HP1100 binary HPLC system. Peptides were eluted with an increasing KCl gradient (10–40 min, 0%–30% buffer B; 40–60 min, 30%–100% buffer B; buffer B consisted of 10 mM KH_2PO_4, pH 2.6, and 350 mM KCl in 25% ACN). The eluate was fractionated into four fractions and desalted with Sep-Pak reverse-phase cartridges (Waters).

IMAC

Chelating Sepharose Fast Flow beads (GE Healthcare) were charged four times with 0.1 M $FeCl_3$ freshly prepared solution and washed four times with washing buffer (74:25:1 water:ACN:acetic acid). Desalted peptides were acidified with 0.1% trifluoroacetic acid (TFA) in 25% ACN, applied to 40 μL of 25% bead slurry, and incubated for 30 min at room temperature. Samples were washed five times with washing buffer and once with water. Phosphopeptides were eluted by adding 30 μL of 100 mM sodium phosphate buffer, pH 8.9. The pH of all samples was adjusted to 3 by adding drops of 10% TFA followed by desalting and concentrating samples using ZipTips (μC18; Millipore).

TiO_2 Affinity Chromatography

Phosphopeptides were enriched using TiO_2 affinity chromatography as described by Bodenmiller et al. (2007) with minor modifications. Peptides were desalted and dissolved in phthalic acid solution (80% ACN, 2.5% TFA, and 0.13 M phthalic acid). The peptide mixture was incubated with 0.3 mg of TiO_2 (GL Science) for 30 min in closed Mobicol spin columns. After washing twice with phthalic acid solution, twice with 80% ACN and 0.1% TFA, once

with 0.1% TFA, and finally again once with 80% ACN and 0.1% TFA, peptides were eluted with 0.3 M NH_4OH and dried in a SpeedVac. Before mass spectrometric analysis, samples were desalted using ZipTips (μC18; Millipore).

Analysis by LC-Electrospray Ionization-MS/MS

Dried peptides were resuspended in 5% ACN and 0.1% formic acid and analyzed on a LTQ-Orbitrap mass spectrometer (ThermoFischer Scientific) interfaced with a nanoelectrospray ion source. Peptides were separated using an Eksigent nano LC system (Eksigent Technologies), equipped with an 11-cm fused silica emitter (75 μm i.d.; BGB Analytik), packed in-house with a Magic C18 AQ 3-μm resin (Michrom BioResources). Peptides were loaded from a cooled (10°C) Spark Holland autosampler and separated using an ACN/water solvent system containing 0.1% formic acid with a flow rate of 200 nL min^{-1}. Peptide mixtures were separated by gradient elution from 3% to 35% ACN in 90 min. Up to five data-dependent MS/MS spectra were acquired in the linear ion trap for each Fourier transform (FT)-MS spectral acquisition range, the latter acquired at 60,000 FWHM nominal resolution settings with an overall cycle time of approximately 1 s. Dynamic exclusion was switched on, ensuring that up to 500 mass-to-charge ratios (m/z) $\pm$ 20 ppm values were excluded from MS/MS for 120 s. For injection control, the automatic gain control was set to $5e^5$ for full FT-MS and to $1e^4$ for linear ion trap MS/MS. The instrument was calibrated externally according to the manufacturer's instructions. The samples were acquired using internal lock mass calibration on m/z 429.088735 and 445.120025.

Data Analysis and MS/MS Spectra Interpretation

MS and MS/MS data were searched against the Arabidopsis database containing common contaminants such as keratin using MASCOT 2.1.04 (Matrix Science) und INSPECT version 14.10.2008 (Payne et al., 2008). Beside carbamylation of Cys residues as a fixed modification, oxidation of Met and phosphorylation of Ser, Thr, and Tyr were included as variable modifications. With MASCOT, database searches were restricted to tryptic peptides, missing maximal two cleavage sites and protein C- and N-terminal peptides, allowing for 2+ and 3+ charged peptides a parent mass error tolerance of 5 ppm and a daughter ion error tolerance of 0.6 D. Identifications were accepted with a MASCOT ion score of 30 or greater and a MASCOT expect value of less than 0.015 (Munton et al., 2007). A normalized delta ion score (ΔI) was calculated for phosphopeptides for which the only difference between the rank 1 and rank 2 hits was the phosphorylation position. ΔI was calculated by taking the difference of the two top ranking ion scores and dividing that difference by the ion score of the first ranking phosphopeptide. Phosphorylation site assignments with $\Delta I \geq 0.4$ were accepted (Elias et al., 2004; Beausoleil et al., 2006). For INSPECT, the search parameters were as follows: instrument = FT-Hybrid (mass tolerance = 100 ppm), maximum of two modifications per peptide, P value cutoff = 0.1. If several hits with a minimum probability of 0.1 were given out for the same spectrum (that only differed in their phosphorylation position), the phosphorylation site assignment was not accepted. Furthermore, none of the Tyr phosphorylation sites that were exclusively based on INSPECT assignments was accepted, because manual data interpretation suggested difficulties of INSPECT with pTyr phosphorylation site assignment (see "Results and Discussion"). All peptide assignments except those of contaminants were filtered for ambiguity, and the peptides matching to several proteins were excluded from further analysis. This does not apply to different splice variants of the same protein or to different loci sharing exactly the same sequence. After database upload, spectrum assignments to decoy database peptides and spectra for which MASCOT and INSPECT assigned to a different peptide were flagged. From the final data, PRIDE 2.1 XML files were created and exported to the PRIDE database. The spectrum false discovery rate was calculated by dividing the number of decoy database spectrum assignments by the number of spectrum assignments in the final data set. For all phosphorylated peptides listed in Table I, further manual spectra inspection was conducted by assigning the parent ion a -98-D neutral loss peak (the loss of phosphoric acid; Supplemental Spectra S1). Relative quantification of the ATPB phosphopeptide was achieved by SuperHirn (Mueller et al., 2007), which aligns and normalizes two LC-MS runs and extracts and processes the area for the identified phosphopeptide peaks. We used SuperHirn version 0.3 from the command line on a Linux computer with the following parameters: MS1 m/z tolerance = 10 ppm, MS1 min tolerance = 1 min. For all other settings, we used default parameters using the MASCOT-generated search files. MzXMLs were generated with readw-4.0.2. Analysis was performed in pairs of LC-MS runs for end of night and the corresponding end-of-day experiment.

GO Classification

Assignment of functional categories was based on the GO categories from aspect "cellular component" (download ATH_GO_GOSLIM.20080510.txt; Berardini et al., 2004). Enrichment analysis of GO categories was done in R (version 2.6.1; http://www.r-project.org) using the "elim" method from the topGO package (version 1.5.1; Alexa et al., 2006) that is part of the Bioconductor project (version 2.1; http://www.bioconductor.org/). Node size was set to five, and Fisher's exact test was used for assessing GO term significance. Overrepresentation of functional categories was calculated for the identified phosphoproteins as compared with the proteins in the TAIR8 (for The Arabidopsis Information Resource) database.

Phosphorylation Motif Analysis Using Motif-X

We applied the Motif-X algorithm (Schwartz and Gygi, 2005) to extract significantly enriched phosphorylation motifs from our phosphopeptide data set. All phosphorylated peptides with confidently identified phosphorylation sites were used as the data set for phosphorylation motif extraction. To focus the motif analysis on the crucial kinase recognition sequence, the peptides were centered at the phosphorylated amino acid and aligned, including six positions upstream and downstream around the phosphorylation site (the necessary sequence context was retrieved from the TAIR8 database). In the case of C- and N-terminal peptides, the sequence was filled up with up to 13 amino acids with the required number of "X." X stands for "any amino acid." The probability threshold was set to $P < 10^{-5}$, and the occurrence threshold was set to 1% of the input data set at a minimum of three peptides. As the background data set, protein sequences of the whole genome Arabidopsis database TAIR8 in Fasta format were used (in a shortened version due to upload restrictions to 10 MB).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Table S1. List of all identified phosphopeptides and phosphoproteins.

Supplemental Table S2. Identified phosphopeptides and phosphoproteins with spectrum assignment information.

Supplemental Table S3. Phosphopeptides of chloroplast proteins.

Supplemental Table S4. Phosphoproteome comparison between different studies.

Supplemental Table S5. TopGO analysis with all identified phosphoproteins.

Supplemental Table S6. Chloroplast proteome reference table assembled from SUBA.

Supplemental Table S7. Identified chloroplast proteins in the FT fractions.

Supplemental Table S8. Phosphopeptides identified at the end of day or the end of night.

Supplemental Table S9. Relative quantification of the ATPB phosphopeptide peaks.

Supplemental Table S10. Protein quantification from the flow-through fractions after IMAC affinity chromatography.

Supplemental Spectra S1. MS/MS spectra for all identified chloroplast phosphopeptides.

ACKNOWLEDGMENTS

We are indebted to Scott Peck and Liliya Serazetdinova for initial advice on phosphoproteome analyses. We also thank Matthias Hirsch-Hoffmann for providing database support.

Received March 13, 2009; accepted April 14, 2009; published April 17, 2009.

LITERATURE CITED

Adams JA (2001) Kinetic and catalytic mechanisms of protein kinases. Chem Rev **101:** 2271–2290

Alexa A, Rahnenfuhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22:** 1600–1607

Baginsky S (2009) Plant proteomics: concepts, applications, and novel strategies for data interpretation. Mass Spectrom Rev **28:** 93–120

Baginsky S, Gruissem W (2002) Endonucleolytic activation directs dark-induced chloroplast mRNA degradation. Nucleic Acids Res **30:** 4527–4533

Baginsky S, Tiller K, Link G (1997) Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (Sinapis alba). Plant Mol Biol **34:** 181–189

Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol **24:** 1285–1292

Bellafiore S, Barneche F, Peltier G, Rochaix JD (2005) State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. Nature **433:** 892–895

Benschop JJ, Mohammed S, O'Flaherty M, Heck AJ, Slijper M, Menke FL (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics **6:** 1198–1214

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, et al (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol **135:** 745–755

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics **4:** 1633–1649

Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods **4:** 231–237

Bollenbach TJ, Schuster G, Stern DB (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. Prog Nucleic Acid Res Mol Biol **78:** 305–337

Bunney TD, van Walraven HS, de Boer AH (2001) 14-3-3 protein is a regulator of the mitochondrial and chloroplast ATP synthase. Proc Natl Acad Sci USA **98:** 4249–4254

de la Fuente van Bentem S, Anrather D, Roitinger E, Djamei A, Hufnagl T, Barta A, Csaszar E, Dohnal I, Lecourieux D, Hirt H (2006) Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism. Nucleic Acids Res **34:** 3267–3278

de la Fuente van Bentem S, Hirt H (2009) Protein tyrosine phosphorylation in plants: more abundant than expected? Trends Plant Sci **14:** 71–76

del Riego G, Casano LM, Martin M, Sabater B (2006) Multiple phosphorylation sites in the beta subunit of thylakoid ATP synthase. Photosynth Res **89:** 11–18

Eberhard S, Finazzi G, Wollman FA (2008) The dynamics of photosynthesis. Annu Rev Genet **42:** 463–515

Elias JE, Gibbons FD, King OD, Roth FP, Gygi SP (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nat Biotechnol **22:** 214–219

Espina V, Edmiston KH, Heiby M, Pierobon M, Sciro M, Merritt B, Banks S, Deng J, VanMeter AJ, Geho DH, et al (2008) A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. Mol Cell Proteomics **7:** 1998–2018

Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA (2007) De novo peptide sequencing and identification with precision mass spectrometry. J Proteome Res **6:** 114–123

Geigenberger P, Kolbe A, Tiessen A (2005) Redox regulation of carbon storage and partitioning in response to light and sugars. J Exp Bot **56:** 1469–1479

Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the Arabidopsis Subcellular Database. Nucleic Acids Res **35:** D213–D218

Jones P, Cote RG, Cho SY, Klie S, Martens L, Quinn AF, Thorneycroft D, Hermjakob H (2008) PRIDE: new developments and new datasets. Nucleic Acids Res **36:** D878–D883

Kanekatsu M, Saito H, Motohashi K, Hisabori T (1998) The beta subunit of chloroplast ATP synthase (CF0CF1-ATPase) is phosphorylated by casein kinase II. Biochem Mol Biol Int **46:** 99–105

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

Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, Baginsky S (2007) Proteome dynamics during plastid differentiation in rice. Plant Physiol **143:** 912–923

Leister D, Kleine T (2008) Towards a comprehensive catalog of chloroplast proteins and their interactions. Cell Res **18:** 1081–1083

Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, Miron IM, Diella F, Colwill K, Taylor L, Elder K, et al (2007) Systematic discovery of in vivo phosphorylation networks. Cell **129:** 1415–1426

Link G (2003) Redox regulation of chloroplast transcription. Antioxid Redox Signal **5:** 79–87

Lisitsky I, Schuster G (1995) Phosphorylation of a chloroplast RNA-binding protein changes its affinity to RNA. Nucleic Acids Res **23:** 2506–2511

Lopez-Juez E, Pyke KA (2005) Plastids unleashed: their development and their integration in plant development. Int J Dev Biol **49:** 557–577

Loza-Tavera H, Vargas-Suarez M, Diaz-Mireles E, Torres-Marquez ME, Gonzalez de la Vara LE, Moreno-Sanchez R, Gruissem W (2006) Phosphorylation of the spinach chloroplast 24 kDa RNA-binding protein (24RNP) increases its binding to petD and psbA 3′ untranslated regions. Biochimie **88:** 1217–1228

Munton RP, Tweedie-Cullen R, Livingstone-Zatchej M, Weinandy F, Waidelich M, Longo D, Gehrig P, Potthast F, Rutishauser D, Gerrits B, et al (2007) Qualitative and quantitative analyses of protein phosphorylation in naive and stimulated mouse synaptosomal preparations. Mol Cell Proteomics **6:** 283–293

Mueller L, Brusniak MY, Mani DR, Aebersold R (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res **7:** 51–61

Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Müller M (2007) SuperHirn: a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics **7:** 3470–3480

Nuhse T, Yu K, Salomon A (2007) Isolation of phosphopeptides by immobilized metal ion affinity chromatography. Curr Protoc Mol Biol **January:** chapter 18, unit 18.13

Nuhse TS, Stensballe A, Jensen ON, Peck SC (2004) Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. Plant Cell **16:** 2394–2405

Ogrzewalla K, Piotrowski M, Reinbothe S, Link G (2002) The plastid transcription kinase from mustard (Sinapis alba L.): a nuclear-encoded CK2-type chloroplast enzyme with redox-sensitive function. Eur J Biochem **269:** 3329–3337

Payne SH, Yau M, Smolka MB, Tanner S, Zhou H, Bafna V (2008) Phosphorylation-specific MS/MS scoring for rapid and accurate phosphoproteome analysis. J Proteome Res **7:** 3373–3381

Pinna LA (1990) Casein kinase 2: an 'eminence grise' in cellular regulation? Biochim Biophys Acta **1054:** 267–284

Preisinger C, von Kriegsheim A, Matallanas D, Kolch W (2008) Proteomics and phosphoproteomics for the mapping of cellular signalling networks. Proteomics **8:** 4402–4415

Puthiyaveetil S, Kavanagh TA, Cain P, Sullivan JA, Newell CA, Gray JC, Robinson C, van der Giezen M, Rogers MB, Allen JF (2008) The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. Proc Natl Acad Sci USA **105:** 10061–10066

Richter ML, Samra HS, He F, Giessel AJ, Kuczera KK (2005) Coupling proton movement to ATP synthesis in the chloroplast ATP synthase. J Bioenerg Biomembr **37:** 467–473

Rochaix JD (2007) Role of thylakoid protein kinases in photosynthetic acclimation. FEBS Lett **581:** 2768–2775

Salih E (2005) Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. Mass Spectrom Rev **24:** 828–846

Salinas P, Fuentes D, Vidal E, Jordana X, Echeverria M, Holuigue L (2006) An extensive survey of CK2 alpha and beta subunits in Arabidopsis: multiple isoforms exhibit differential subcellular localization. Plant Cell Physiol **47:** 1295–1308

Schliebner I, Pribil M, Zuhlke J, Dietzmann A, Leister D (2008) A survey of chloroplast protein kinases and phosphatases in Arabidopsis thaliana. Curr Genomics **9:** 184–190

Schwartz D, Gygi SP (2005) An iterative statistical approach to the

identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol **23:** 1391–1398

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol Syst Biol **4:** 193

Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem **77:** 4626–4639

Tiller K, Link G (1993) Phosphorylation and dephosphorylation affect functional characteristics of chloroplast and etioplast transcription systems from mustard (Sinapis alba L.). EMBO J **12:** 1745–1753

Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci USA **104:** 1488–1493

Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem **138:** 141–143