Phosphoproteomics of the Arabidopsis Plasma Membrane and a New Phosphorylation Site Database [W]

Thomas S. Nühse,[a] Allan Stensballe,[b] Ole N. Jensen,[b] and Scott C. Peck[a,1]

[a] Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, United Kingdom

[b] Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark

Functional genomic technologies are generating vast amounts of data describing the presence of transcripts or proteins in plant cells. Together with classical genetics, these approaches broaden our understanding of the gene products required for specific responses. Looking to the future, the focus of research must shift to the dynamic aspects of biology: molecular mechanisms of function and regulation. Phosphorylation is a key regulatory factor in all aspects of plant biology; but it is difficult, if not impossible, for most researchers to identify in vivo phosphorylation sites within their proteins of interest. We have developed a large-scale strategy for the isolation of phosphopeptides and identification by mass spectrometry (Nühse et al., 2003b). Here, we describe the identification of more than 300 phosphorylation sites from *Arabidopsis thaliana* plasma membrane proteins. These data will be a valuable resource for many fields of plant biology and overcome a major impediment to the elucidation of signal transduction pathways. We present an analysis of the characteristics of phosphorylation sites, their conservation among orthologs and paralogs, and the existence of putative motifs surrounding the sites. These analyses yield general principles for predicting other phosphorylation sites in plants and provide indications of specificity determinants for responsible kinases. In addition, more than 50 sites were mapped on receptor-like kinases and revealed an unexpected complexity of regulation. Finally, the data also provide empirical evidence on the topology of transmembrane proteins. This information indicates that prediction programs incorrectly identified the cytosolic portion of the protein in 25% of the transmembrane proteins found in this study. All data are deposited in a new searchable database for plant phosphorylation sites maintained by PlantsP (http://plantsp.sdsc.edu) that will be updated as the project expands to encompass additional tissues and organelles.

INTRODUCTION

After the completion of several eukaryotic genome sequences and with the automation of gene expression studies, biologists face the prospect that the next step toward understanding cellular processes is far more complex. The functions of proteins are regulated by localization, binding partners, posttranslational modifications, and stability. Most of these parameters are poorly predictable from their primary sequence—a postgenomic challenge. It is thus necessary to examine the protein itself, which can be a formidable task for regulatory proteins of low abundance. In the past few years, rapid progress in mass spectrometric technology has allowed analysis and identification of proteins with unprecedented levels of sensitivity, and numerous proteomic studies have tackled the aforementioned challenges. The protein complements of subcellular structures have been explored by organellar proteomics (Taylor et al., 2003). The interaction partners of hundreds of yeast proteins identified with affinity purification techniques (Gavin et al., 2002; Ho et al., 2002) and the sites of posttranslational modifications were determined using specific enrichment techniques (Ficarro et al., 2002; Elortza et al., 2003; Peng et al., 2003).

Protein phosphorylation is one of the most widespread and arguably best understood posttranslational modifications. Virtually all cellular processes are regulated in one or multiple ways by phosphorylation and dephosphorylation, and the identification of kinases, their substrates, and the specific site of phosphorylation is the key to a molecular understanding of signaling. In vivo labeling (Peck et al., 2001), phosphorylation-specific antibodies, and/or band shifts in immunoblots are all methods to show that a particular protein becomes phosphorylated in response to a stimulus. The identification of the actual site(s) of phosphorylation, however, is necessary for a functional characterization and represents a major bottleneck for analysis of individual proteins. With the described advances in mass spectrometric technology, we can now approach the problem from a different angle. Phosphopeptides can be isolated from complex protein digests and identified in a large scale by mass spectrometry in a shotgun approach (Ficarro et al., 2002). In a sufficiently large study, the phosphorylation sites of interest may be found directly. Alternatively, the identified sites from related proteins provide enough knowledge to guide further experiments.

We know from the *Arabidopsis thaliana* genome that plants have evolved signal transduction components that are substantially different from those of animals, in particular Ser/Thr rather

than Tyr receptor kinases (Arabidopsis Genome Initiative, 2000). Uncritical transfer of mammalian signaling paradigms may therefore be misleading, and it will be necessary to establish mechanisms that are unique in plants. Two major obstacles, however, impede the analysis of integral membrane signaling proteins. First, the proteins are difficult to work with biochemically because of their hydrophobic nature and typically low abundance. Second, identification of phosphorylation sites requires an enrichment method for phosphopeptides because suppression effects during ionization prevent the detection of these peptides by mass spectrometry when in complex mixtures.

The problems of membrane protein insolubility are circumvented by proteolytic digestion of the intact membranes and analysis of peptides released from extramembrane domains (described in Nühse et al., 2003b). Because the proteins are digested in situ and only the fragments from intracellular loops and N and C termini recovered, no selection against large proteins or those with multiple transmembrane domains is apparent. This strategy, then, overcomes the first limitation.

For enrichment of phosphopeptides, immobilized metal-ion affinity chromatography (IMAC) for phosphopeptides showed great promise for large-scale studies but had a reputation for poor specificity. We investigated the potential of IMAC in combination with capillary liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for the identification of plasma membrane phosphoproteins of Arabidopsis (Nühse et al., 2003b). Using either simple batch binding of peptides or prefractionation via strong anion exchange (SAX) chromatography before IMAC, a distinct set of phosphopeptides was detected. The degree of enrichment over nonphosphorylated peptides (75 to 90% purity) allowed efficient analysis and solved the second challenge.

Here, we present an analysis of several hundred phosphorylation sites on ∼200 putative plasma membrane proteins of Arabidopsis collected from multiple LC-MS/MS analyses of independent plasma membrane preparations. This phosphoproteome study is among the largest performed in any organism. Besides providing new insights into the functional regulation of numerous proteins, the size of the data set enabled us to find characteristics and patterns that form the basis for developing predictive models of phosphorylation in plants.

RESULTS

Overcoming Limitations in Plasma Membrane Phosphoproteome Analysis

Two benchmark parameters were used to test whether the methodology delivers comprehensive proteome coverage: the representation of proteins with multiple transmembrane-spanning domains and the sequence coverage of cytoplasmic domains that can be phosphorylated. Analyzing the data from this large-scale phosphoproteomic study, we found that approximately two-thirds of the identified phosphoproteins had at least one transmembrane (TM) domain (Figure 1A). One of them, a calpain-like protein, contained a predicted 23 TM domains. It has been suggested that precleavage of membrane proteins with
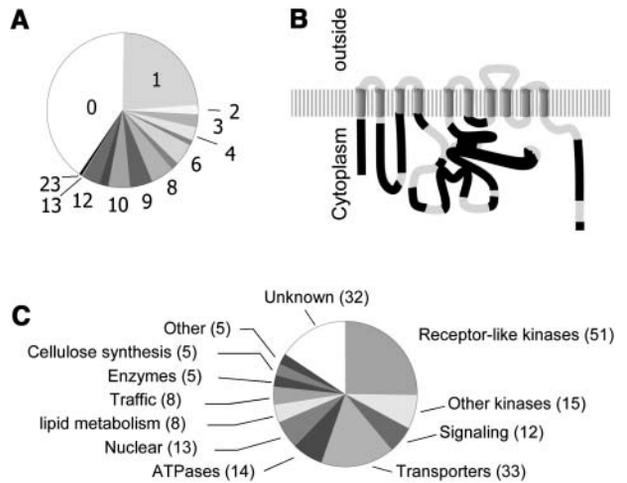

Figure 1. Coverage and Functional Classification of PM Phosphoproteins.

(A) Proportional representation of transmembrane helices within proteins found in this study as predicted by TMHMM (http://www.cbs.dtu.dk/services/TMHMM/).
(B) Graphical representation of sequence coverage of AHA2 in a shotgun proteomic experiment with total PM peptides. Structural model adapted from Palmgren (2001).
(C) Representation of functional protein classes among PM phosphoproteins.

cyanogen bromide is necessary for complete tryptic digestion (Washburn et al., 2001). In an analysis of a total plasma membrane (PM) digest, however, we found very good coverage of even short cytoplasmic loops for the proton ATPase AHA2 (Figure 1B). We postulate that membrane shaving of tryptic peptides is facilitated by the fact that TM domains are often anchored in the membrane by polybasic residues (i.e., trypsin cleavage sites) on the cytosolic face. Thus, the accessibility of PM proteins to trypsin does not appear to be a limiting factor. Moreover, all fragments were from regions predicted to reside on the cytoplasmic face of the PM (Figure 1B), indicating that this type of analysis provides information on the topography of proteins within the membrane (Wu et al., 2003).

The functional categories of the 200 identified phosphoproteins (Nühse et al., 2003b; this study) show a high representation of proteins with a role in signal transduction. Receptor-like kinases represent a quarter of all phosphoproteins and all signaling-related proteins, as far as could be annotated by similarity, more than a third. Proteins with a role in transport (channels and ABC transporters) and ion homeostasis (H^+ and Ca^{2+} ATPases) constitute another large family. Although the plasma membranes were of high purity (Nühse et al., 2003a), a small number of proteins were identified as clear contaminants, among them several nuclear proteins, such as DNA polymerases and putative transcription factors.

Confirmed and Novel Phosphorylation Sites

Unsurprisingly, all but a few phosphorylation sites described here are novel. Among the few well-characterized plant plasma

membrane phosphoproteins are the H^+-ATPases and the aquaporins. We reported the identification of previously known and novel phosphorylation sites on several isoforms of the H^+-ATPase (AHA1, AHA2, AHA3, and AHA4) in an earlier publication (Nühse et al., 2003b). Vacuolar and plasma membrane aquaporins are known to be regulated by posttranslational modification to adapt to rapid changes in the apoplastic water potential (Johansson et al., 1996; Maurel, 1997). The phosphorylation sites we identified on two PIP2 isoforms (see Supplemental Table 1 online) have also been detected in another proteomic study (Santoni et al., 2003). They are equivalent to Ser262 of soybean (*Glycine max*) nodulin26 and Ser274 of spinach (*Spinacia oleracea*) PM28A (Johansson et al., 1998; Guenther et al., 2003), both of which are phosphorylated in response to increasing water potential, probably by a calcium-dependent kinase.

In several cases, phosphorylation of a particular protein was known or predicted, but no data on the precise site of modification were published. The Na^+/H^+-antiporter SOS1 is phosphorylated by the SOS2 kinase in complex with the calcium sensor protein SOS3 on (an) unknown site(s) (Quintero et al., 2002). The site we identified is consistent with both the anticipated SNF1/AMPK1 motif and the localization on the hydrophilic C terminus (Halfter et al., 2000), and thus may be (one of) the authentic site(s) targeted by SOS2/SOS3.

For the remainder of the phosphorylation sites, however, this report represents the first insights into potential posttranslational regulation of these proteins. In addition to the importance in studying individual proteins, the data provide the basis of drawing some general predictions and conclusions about protein phosphorylation in plants.

Phosphorylation Sites Tend to Cluster Outside Known Functional Domains

Proteins are generally composed of catalytic/functional domains surrounded or interspersed by less defined (or not yet understood) variable regions. We analyzed our data to investigate if there is a trend in the location of phosphorylation sites in relation to the architecture of the protein using protein domain predictions from the SMART Web site (http://smart.embl-heidelberg.de/).

Figure 2 shows the position of phosphorylation sites relative to functional domains for three multidomain protein families represented in our study. In the majority of cases, the sites were outside catalytic and known regulatory domains, and sequences surrounding the sites were poorly conserved within a gene family. Nevertheless, such isoform-specific regulatory sites can still be clustered in equivalent regions of the protein. In the case of the multiple drug resistance (MDR) family of ABC transporters, the poorly conserved linker region between the duplicated ATPase domains is a well-known target for phosphorylation in yeast and human transporters (Decottignies et al., 1999; Idriss et al., 2000). Three sites in plant ABC transporters share an [R,K]-X-pS motif (see Supplemental Table 1 online), suggesting that these sites are phosphorylated by the same or a related kinase. In the case of the cellulose synthase family, all but one of the sites are in the unconserved region upstream of the first TM domain, whereas one phosphorylation motif in the catalytic domain of
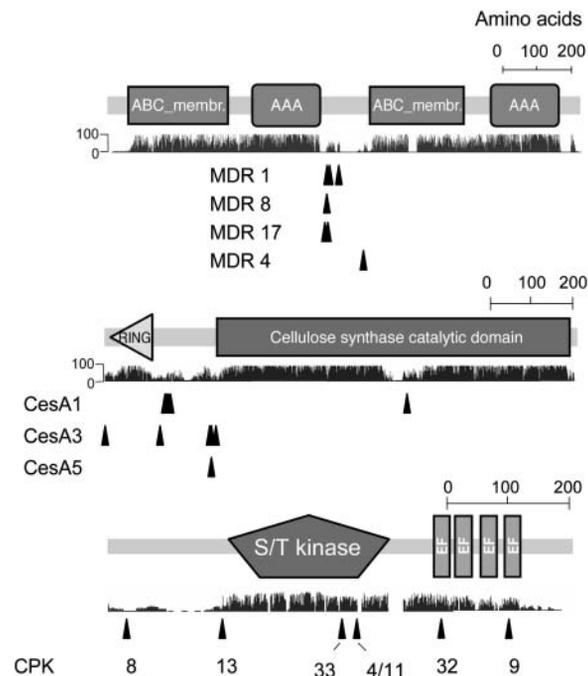

Figure 2. Graphical Representation of Protein Architecture, Sequence Conservation within the Protein Family, and Localization of Phosphorylation Sites.

Arrowheads indicate the location of phosphorylation sites within the protein. Protein domain architecture was adapted from the SMART Web site (http://smart.embl-heidelberg.de/). Sequence conservation was plotted from ClustalW alignments using MDR family members as listed in PlantsT (http://plantst.sdsc.edu/) and CDPK members from PlantsP (http://plantsp.sdsc.edu/). Arabidopsis CesA sequences were downloaded from http://cellwall.stanford.edu/php/download.php. ABC_membr., TM domain of ABC transporters; AAA, ATPases associated with various cellular activities; RING, RING-type zinc finger; EF, EF hand/calcium binding motif.

CesA1 is loosely conserved among several other CesA proteins. Among the calcium-dependent protein kinases, we found two sites within the C-terminal calmodulin-like domains (where in one case the phosphoserine replaces a conserved Asp, see below), two in the kinase domain, and only two in the highly diverse N termini. Taking into account the data from receptor-like kinases (see below) and other families (data not shown), the majority of phosphorylation sites are located outside of defined catalytic/functional domains.

An additional outcome of this analysis was that phosphorylation sites in most of the multispanning TM proteins were close to either the N or C terminus. Assuming that phosphorylation is exclusively intracellular, the identification of a phosphopeptide establishes the corresponding terminus as cytoplasmic. This knowledge is useful for the determination of the inside/outside topology of unknown membrane proteins. We found a Web-based prediction tool (http://www.cbs.dtu.dk/services/TMHMM/) in conflict with our empirical data for more than one-quarter of proteins with more than one TM domain (data not

shown). Thus, these data have significance on a structural as well as functional level.

Conservation of Phosphorylation Sites between Related Proteins

Related to the question addressed above, we also investigate if the amino acid sequence surrounding a phosphorylation site is conserved between members of functional families of proteins. It could be envisaged that sites involved in the direct regulation of enzyme activity might be more likely to occur within conserved regions of a protein family. On the other hand, phosphorylation of unique sites would allow isoform-specific regulation. In the latter case, these sites may still be conserved among true orthologs in related plant species.

Figure 3 shows an alignment of the N-terminal phosphorylation sites of the cellulose synthase CesA3 with the corresponding sequences of paralogous CesA proteins (top panel) and putative orthologs (bottom panel), the latter found by BLAST search of EST databases. It is striking that whereas the CesA3 sites are unique among the paralogous CesA proteins, highly similar sequences were found in putative orthologs from the EST database, indicating that isoform-specific regulation of this cellulose synthase protein is conserved even in monocots.

To get a more representative overview of sequence conservation around phosphorylation sites, we did a similar comparison for all putative transporter proteins. Putative paralogs were taken from the Aramemnon database (http://aramemnon.botanik. uni-koeln.de/). Putative orthologs were found by BLAST search of ~40 to 60 amino acids surrounding the phosphorylation site against an EST database, including a few amino acids of sequence that are conserved in the protein family when possible to reduce random low-scoring hits. As shown in Table 1, the majority of phosphorylation sites are conserved among putative orthologs and, to a lesser extent, among some members of the same protein family. Because of potential errors in EST sequence data, a negative result cannot be taken as evidence for absence of related phosphorylation sites. Conversely, however, the conservation of a putative phosphorylation motif may be taken as additional supporting evidence for the exact phosphorylation site if the mass spectrometric data are ambiguous. The SOS1 phosphorylation site could not be unambiguously assigned to S1 or S3 in the peptide SVSFGGIYNNK, but S3 is more highly conserved in other species. Thus, it is the more likely candidate (Figure 4A). In addition, careful analysis of aligned sequences may also help to define specificity determinants of the kinase that targets the sequence. In SOS1, the Phe residue in the +1 position is also conserved and may be important in addition to the known
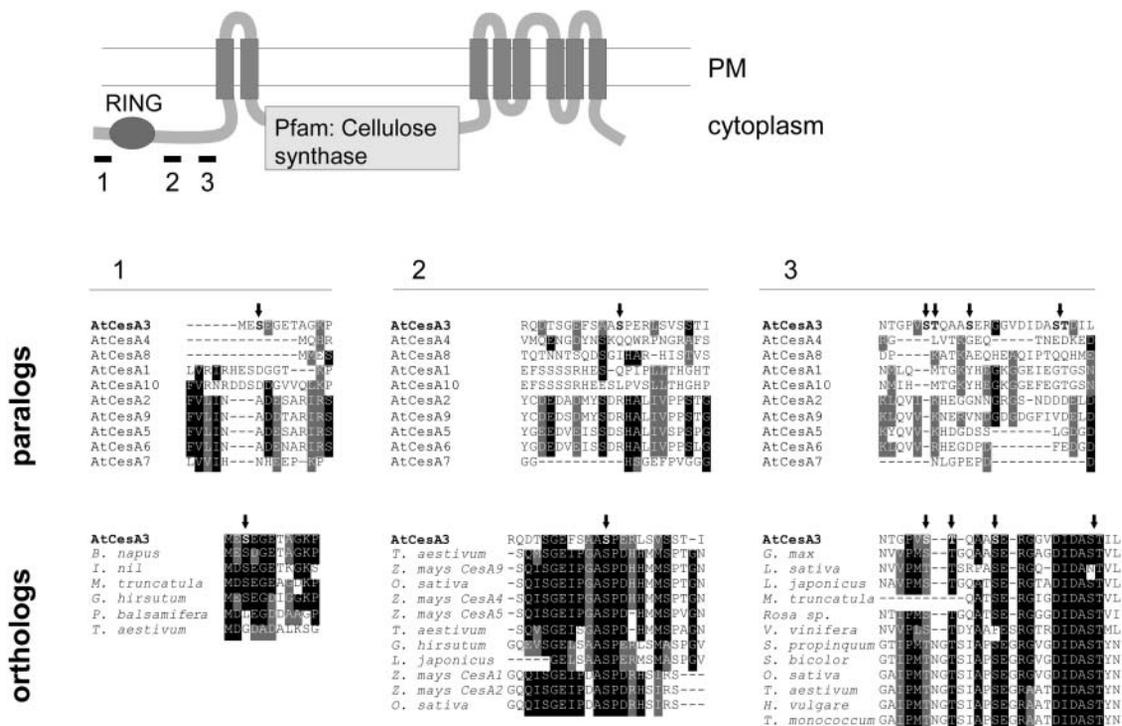
Figure 3. Comparison of Phosphorylation Sites Conserved in Paralogs and Orthologs of Cellulose Synthases.

Three regions surrounding CesA3 phosphorylation sites (marked with arrows) have been aligned with all paralogous CesA sequences from Arabidopsis (top panel) and with putative orthologs in other species. The orthologs were found by BLAST search in EST databases using 40 to 60 amino acids from CesA3. The location of the short sequences relative to the functional domains is indicated above the alignments. Identical residues are in black boxes, and similar residues are in gray.

Table 1. Conservation of Phosphorylation Sites in Membrane Transporters

PlantsT	Family	Subfamily/Name	Protein	Phosphorylation Site	Conservation within Protein Family	Conservation among Putative Orthologs (EST Database)
2.A.1	STP	Monosaccharide transporter	STP1	RGVDDV**pS**QEFDDL	–	(+)
			STP13	FVNGEK**pS**NGKSNG	–	+
	PHT	Phosphate transporter	PHT1-4	NEDNEN**pS**NNDSRT	–	n.f.
2.A.2	SUC	Sucrose-proton symporter	SUC5	AANNApTALEpTQpSpSPEDLGQ	3 (9)	+
2.A.7	PUP	Purine permease	PUP18	EPDQIL**pS**PRRSLE	2 (13)	+
2.A.16	TDT	C4-dicarboxylate/malic acid transporter (telluride resistance)	At4g13800	PVFINpSGpSSRSSNS	2 (7)	n.f.
			At5g24030	DQLRNVpSpSENIENYLK	–	n.f.
2.A.17	POT	Proton-dependent oligopeptide transporter	At3g47960	NTDVVD**pS**FEEEQR	–	+
			At1g69870	GSFSKSpSpSPpSELDV.. YKRIpSpSPGSILD	–	(+)
2.A.18	AAAT	amino acid/auxin permease	AtLTH4	PATPRV**pSpT**PEILTP	–	+
			At5g41800	PVTRLD**pS**DAGALF	–	(+)
			At5g01240	GRKVED**pS**AAEEDI	–	–
2.A.29		Mitochondrial carrier	At3g55640	KFMYMV**pT**GMENHK	–	+
2.A.31	AE	Anion exchangers/boron transporter	At3g06450	EAPAIL**pS**FNLKPE	–	+
2.A.36	NHX	Sodium-proton exchanger	SOS1/NHX7	VNVYRR**pS**VSFGGI	–	+
			NHX1,2	VPFVPGpSPpTERNPPD	–	+
2.A.40	NCS2	Nucleobase:cation symporter 2	At2g27810	GFRPKFpSGEpTTApTDpSpSSGQLSL	–	+
2.A.49	AMT	Ammonium transporter	AMT1	MAGMDM**pT**RHGGFA	3 (6)	+
2.A.55	NRAMP	Metal ion (Fe/Mn) transporter	NRAMP1	QLPCRV**pS**TSDVD	–	+
2.A.57	ENT	Equilibrative nucleoside transporter	ENT6	AGIQNQ**pS**DL**pS**DDDSKN	4 (7)	+
2.A.72	KUP	K⁺ uptake permease	KUP2	EDDNAR**pS**VQ**pS**NE**pS**SSESR	–	(+)
			KUP11	DQKLDQ**pS**MDEEAG	–	–

The conservation of a phosphorylatable residue (S or T) in paralogs and putative orthologs of phosphorylated transporters was analyzed. Phosphoproteins were grouped into families according to PlantsT (http://plantst.sdsc.edu/). Proteins listed in the Aramemnon database (http://aramemnon.botanik.uni-koeln.de/) as isospecific homologs of the phosphoprotein with at least 30% similarity were called protein family. Under the heading Conservation within Protein Family, the number of family members that share the phosphorylation site is listed. Under the heading Conservation among Putative Orthologs, (+) denotes one BLAST hit in the EST database with conserved phosphorylation site; + denotes two or more hits; n.f., no ambiguous ortholog was found in the EST databases.

basic-X-X-Ser motif for the SNF1/AMP-related kinase SOS2 (Weekes et al., 1993; Qiu et al., 2002).

In the glucose transporter STP1 (Figure 4B) and, most strikingly, the calcium-dependent protein kinase CPK32 (Figure 4C), we found the interesting case of a conserved acidic residue being replaced by a phosphoamino acid in one or two sequences. In CPK32, the phosphoserine takes the place of an absolutely conserved Asp that coordinates the calcium ion in the EF hand. It is tempting to speculate that phosphorylation of this Ser residue represents the switching on of a charge-dependent function in one isoform where the others are constitutively on.

Phosphorylation of Receptor-Like Kinases Implies Complex Regulation

Receptor-like Ser/Thr kinases (RLKs) form an extremely diverse protein family in Arabidopsis. Because the genome sequence has revealed their number to be in the hundreds, the question has arisen of how signaling specificity is ensured. Even if only a fraction of all RLK genes is expressed at any given time in

one cell, it is puzzling how unique output signals are generated from receptors with highly similar kinase domains.

Our experimental data on RLK phosphorylation sites are perhaps the most striking of this study. With ~50 identified phosphoproteins, this protein family is by far the best represented and accounts for a quarter of all proteins (Figure 1C). Figure 5A shows a graphic representation of the position of the phosphorylation sites relative to the conserved kinase subdomains (for a full-size alignment, see Supplemental Figure 1 online). Unexpectedly, three-quarters of the phosphopeptides came from either the juxtamembrane region ([JM]; the linker between the TM domain and the kinase core) or the domain C-terminal of the kinase core. The JM and C-terminal sequences are highly divergent in length and sequence, and the identified phosphorylation sites are generally unique for a single RLK (data not shown). The implications for signaling mechanisms are discussed below.

Only eight phosphopeptides were found in the activation segment or A loop (i.e., the sequence between the conserved DFG and APE motifs of the subdomains VII and VIII, respectively),
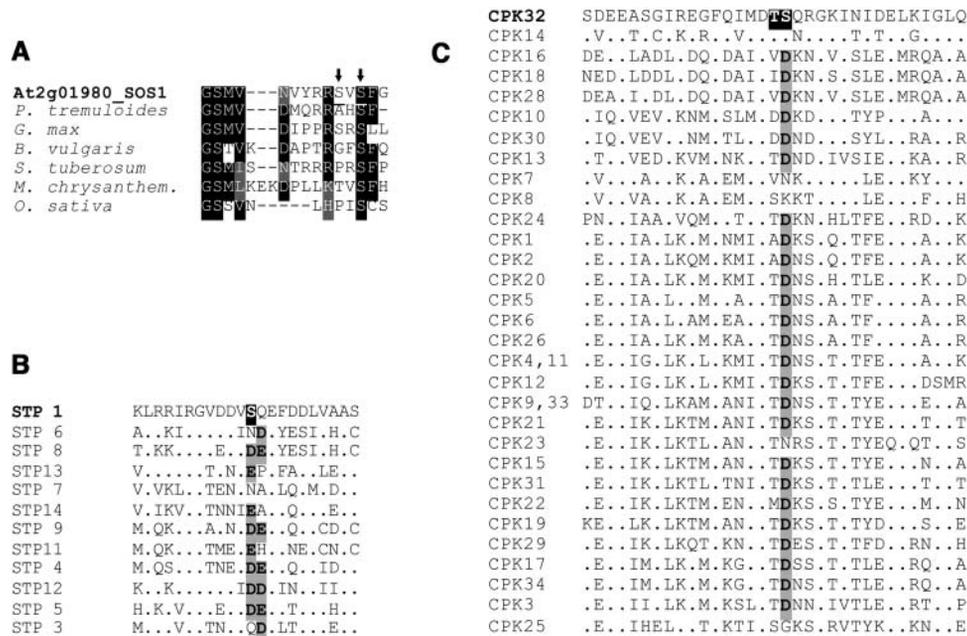
A

```
At2g01980_SOS1   GSMV---NVYRRSVSFG
P. tremuloides   GSMV---DMQRRAHSH-
G. max           GSMV---DIPPRSRSLL
B. vulgaris      GSIVK--DAPTRGFSEQ
S. tuberosum     GSMIS--NTRRRPRSPP
M. chrysanthem.  GSMLEKDPLLKTVSEH
O. sativa        GSSVN-----LEPISCS
```

B

```
STP 1    KLRRIRGVDDVSQEFDDLVAAS
STP 6    A..KI.....IND.YESI.H.C
STP 8    T.KK....E..DE.YESI.H.C
STP13    V......T.N.EP.FA..LE..
STP 7    V.VKL..TEN.NA.LQ.M.D..
STP14    V.IKV..TNNIEA..Q..E...
STP 9    M.QK...A.N.DE..Q..CD.C
STP11    M.QK...TME.EH..NE.CN.C
STP 4    M.QS...TNE.EH..Q..ID..
STP12    K..K......IDD.IN..II..
STP 5    H.K.V..E..DE..T...H..
STP 3    M...V..TN..QD.LT...E..
```

C

```
CPK32    SDEEASGIREGFQIMDTSQRGKINIDELKIGLQ
CPK14    .V..T.C.K.R..V.....N.....T.T..G....
CPK16    DE..LADL.DQ.DAI.VDKN.V.SLE.MRQA.A
CPK18    NED.LDDL.DQ.DAI.IDKN.S.SLE.MRQA.A
CPK28    DEA.I.DL.DQ.DAI.VDKN.V.SLE.MRQA.A
CPK10    .IQ.VEV.KNM.SLM.DDKD...TYP...A...
CPK30    .IQ.VEV..NM.TL..DDND...SYL..RA..R
CPK13    .T..VED.KVM.NK..TDND.IVSIE..KA..R
CPK7     .V...A..K.A.EM..VNK.....LE..KY...
CPK8     .V..VA..K.A.EM..SKKT....LE...F..H
CPK24    PN..IAA.VQM..T..TDKN.HLTFE..RD..K
CPK1     .E..IA.LK.M.NMI.ADKS.Q.TFE...A..K
CPK2     .E..IA.LKQM.KMI.ADNS.Q.TFE...A..K
CPK20    .E..IA.LK.M.KMI.TDNS.H.TLE...K..D
CPK5     .E..IA.L..M..A..TDNS.A.TF....A..R
CPK6     .E..IA.L.AM.EA..TDNS.A.TF....A..R
CPK26    .E..IA.LK.M.KA..TDNS.A.TF....A..R
CPK4,11  .E..IG.LK.L.KMI.TDNS.T.TFE....A..K
CPK12    .E..IG.LK.L.KMI.TDKS.T.TFE...DSMR
CPK9,33  DT..IQ.LKAM.ANI.TDNS.T.TYE...E..A
CPK21    .E..IK.LKTM.ANI.TDKS.T.TYE...T..T
CPK23    .E..IK.LKTL.AN..TNRS.T.TYEQ.QT..S
CPK15    .E..IK.LKTM.AN..TDKS.T.TYE...N..A
CPK31    .E..IK.LKTL.TNI.TDKS.T.TLE...T..T
CPK22    .E..IK.LKTM.KN..MDKS.S.TYE...M..N
CPK19    KE..LK.LKTM.AN..TDKS.T.TYD...S..E
CPK29    .E..IK.LKQT.KN..TDES.T.TFD..RN..H
CPK17    .E..IM.LK.M.KG..TDSS.T.TLE..RQ..A
CPK34    .E..IM.LK.M.KG..TDNS.T.TLE..RQ..A
CPK3     .E..II.LK.M.KSL.TDNN.IVTLE..RT..P
CPK25    .E..IHEL..T.KTI.SGKS.RVTYK..KN..E
```

Figure 4. Phosphorylation of Ser Residues May Act as a Reversible Mimic of Functionally Important Acidic Residues.

(A) ClustalW alignment of the experimentally determined phosphorylation site on the Na+/H+ transporter SOS1 with putative orthologous sequences. Potential phosphorylation sites in SOS1 are indicated by arrows.

(B) and **(C)** Alignment of sequences corresponding to phosphorylation sites (white on black background) in the glucose transporter STP1 and the calcium-dependent protein kinase CPK32, respectively. Acidic residues in the position equivalent to experimentally determined phosphorylation sites have a gray background.

the best-characterized domain for kinase activation. Several of them cover the same conserved residues, and one of them (GpSFGYLD...) corresponds to one of the very few previously known RLK phosphorylation sites, Thr468 in SERKa (Shah et al., 2001). Most, but not all, RLKs with phosphorylated A loops belong to the group of RD kinases that require this modification for charge neutralization in the active site (Johnson et al., 1996).

Outside of the activation segment, some more unusual phosphorylation sites were found. Several peptides cover an [RK]-A-S-A-E motif just upstream of domain I, which is highly conserved among the LRRK III and VII families and probably represents a common regulatory site. Modification of residues in the unconserved linker regions between the conserved subdomains I/II, II/III, and IX/X, respectively, or within domains X/XI (Figure 5) have to our knowledge not been reported for any protein kinase. This certainly reflects a bias in hypothesis-driven research toward previously known sites (particularly in the A loop), but these unusual sites may indeed prove to be unique for plant RLKs or even the particular protein itself.

Phosphorylation Site Motifs

Although kinases may have additional protein–protein docking or interaction domains, the enzymatic active site shows specificity toward its substrates based on the linear sequence surrounding the residue to be phosphorylated. We sorted all unambiguously assigned phosphorylation sites according to single residues in the −6 to +6 positions to detect the frequency of residues within specific positions and to look for common motifs (Figure 6A). Although most residues were evenly distributed across all positions, Pro was seldom in the −1 position and overrepresented in the +1. Aromatic residues were also found with disproportionate frequency in the +1 position but were virtually excluded from position +2. Taking these analyses further, we found restraint on the residues in the −1 position when an aromatic residue was in the +1 position. Overall, ~12% of total Gly residues are present in the −1 position. However, when an aromatic residue is present in the +1 position, Gly is the −1 residue 50% of the time (13/26; Figure 6B). Conversely, Pro or aromatic residues are never present at the −1 position with an aromatic residue in the +1 position. These indications of bias probably reflect spatial restraints of the substrate binding site of the relevant kinases. Information from these types of analyses will be useful both for designing peptide libraries to test kinase specificity determinants and for refining programs for predicting phosphorylation sites.

Finally, we investigated the possibility of using patterns of residues surrounding the phosphorylation site to predict kinase specificity determinants or to cluster proteins that may be coordinately regulated. We found an unusually large number of motifs with acidic residues in the −1 to +4 positions (see Supplemental Figure 2D online), many even with a [DE]-pS-[DE] motif (see Supplemental Figure 2E online). The second largest group were protein kinase A-like motifs [RK]-x_{1-3}-pS (see Supplemental Figures 2A and 2B online). The high frequency of
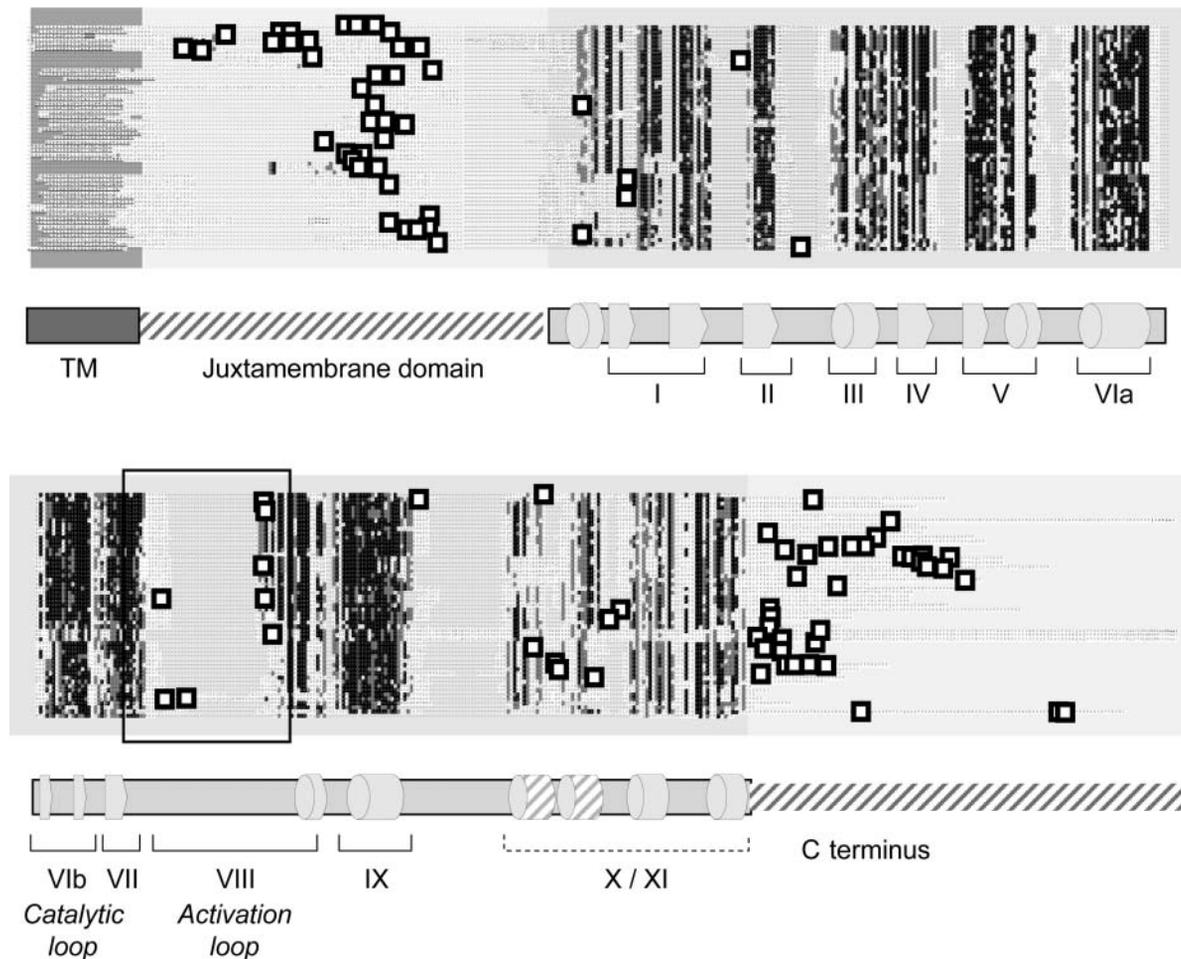
Figure 5. Phosphorylation Sites of Receptor-Like Kinases.

A ClustalW alignment of phosphorylated receptor-like kinases was rendered with Boxshade. Experimentally determined phosphorylation sites are represented by white squares with black frames. A schematic drawing of the secondary structure (below) shows conserved α helices and β sheets of the kinase domain as cylinders and arrows, respectively. An enlargement of this sequence alignment is provided in Supplemental Figure 1 online.

these motifs reflects the fact that acidic- or basic-directed specificity is common among the well-characterized kinases. The G – pS – aromatic motif discussed above is conserved among most PIP2 (plasma membrane intrinsic protein) and all Syp1 (syntaxin family 1) proteins (Figure 6A). Interestingly, phosphorylation of the homologous site of the aquaporin PM28A has been shown to be calcium dependent (Weaver and Roberts, 1992), and uncharacterized calcium-dependent activities also phosphorylated a H^+-ATPase (Schaller and Sussman, 1988) and the syntaxin Syp122 (Nühse et al., 2003a). It is conceivable that this unusual phosphorylation motif is subject to regulation by the same or closely related kinases. If these kinases are indeed CDPKs, the substrate motifs are more variable than anticipated (Cheng et al., 2002). Several motifs have been described as preferred substrates for the related Snf1-like and calcium-dependent protein kinases (Weekes et al., 1993; McMichael et al., 1995; Bachmann et al., 1996; Huber and Huber, 1996; Huang and Huber, 2001; Huang et al., 2001). Many of the experimentally determined sites correspond to the simple

pattern basic-X_2-pSer, but only very few match the ideal motif determined in peptide-based assays (see Supplemental Figure 2 online). It is possible that the plasma membrane samples contain only very few substrates of Snf1-like kinases, such as metabolic enzymes, but a more likely explanation is that full-length proteins in vivo reveal a more complex recognition motif than a short stretch of primary structure. Thus, the amino acid positions described here for sorting may not actually be the sole specificity determinant for the respective kinase. It is likely, however, that the inclusion of structural and other parameters as well as empirical data on kinase specificity will allow us to use these data sets for predictions as was successfully done for mammalian proteins (Blom et al., 1999).

DISCUSSION

With the set of Arabidopsis plasma membrane phosphorylation sites presented in this work, we have made two major achievements. First, as a database resource, the data will facilitate
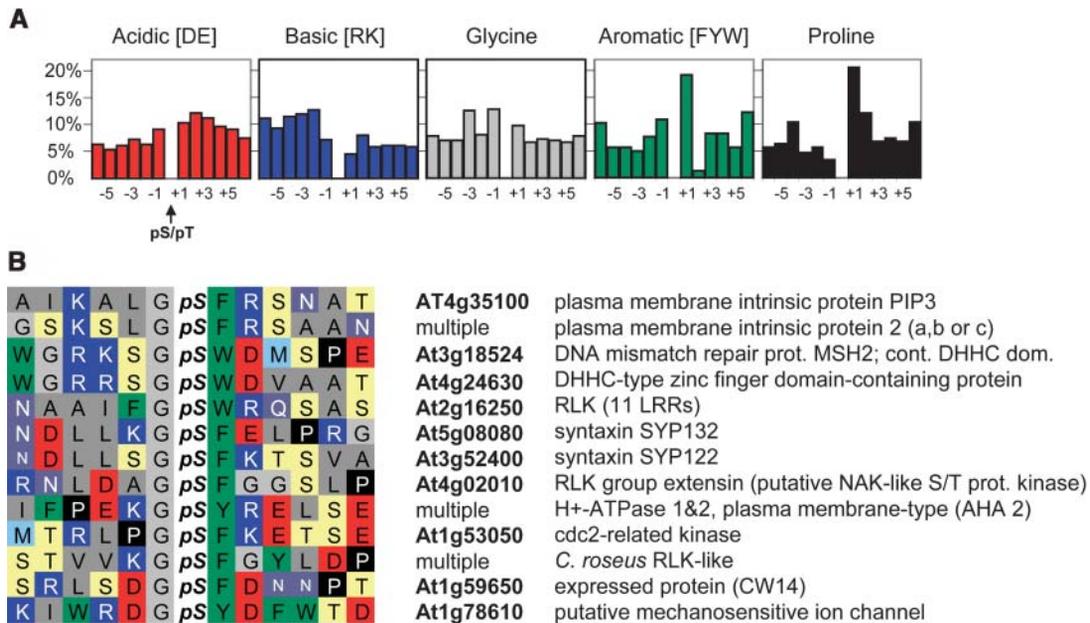
Figure 6. Frequency Distribution of Amino Acid Residues Surrounding Phosphorylation Sites.

(A) Frequency of amino acids in single positions relative to the total frequency in positions −6 to +6.
(B) Alignment of phosphorylation site sequences (residues −6 to +6) with the experimentally found G-pS-[FYW] motif. The color scheme is simplified from the scheme in (Taylor, 1997).

signaling research on the proteins identified. Technically, it remains difficult, if not impossible, for most laboratories to identify phosphorylation sites within their proteins of interest. These data, and those that follow, should accelerate studies into the mechanistic regulation of biological processes in plants in a manner comparable to the way genome and EST sequencing projects have facilitated the cloning of genes of interest. All data will be released as a free access, searchable database through PlantsP (http://plantsp.sdsc.edu). Second, in a wider scope, the data provide insight into general emerging patterns of regulation through phosphorylation as well as an appreciation of the complexity of these processes.

It is important to stress that the identification of phosphorylation sites as described in this work relies on MS/MS spectra from a single peptide. Thus, the analysis requires a sequenced genome to ensure a high-confidence assignment and currently restricts such large-scale plant studies to Arabidopsis and rice (*Oryza sativa*). We found, however, that phosphorylation sites identified in Arabidopsis generally are conserved in a broad range of species (Table 1, Figures 3 and 4; data not shown). These results indicate that the evolutionary conservation of regulatory mechanisms allows for the direct transfer of phosphorylation site data from model species to crop species.

Concerning individual phosphorylation sites, our key findings are posttranslational regulation of many novel proteins and protein families and a comprehensive insight into the complexity of receptor-like kinase phosphorylation. For most of the identified protein families, such as receptor-like kinases, ABC transporters, and ion channels, posttranslational regulation in yeast and mammals is well characterized. Therefore, finding

phosphorylation sites on analogous Arabidopsis proteins was not surprising. For the family of cellulose synthases and the functionally associated cellulase KORRIGAN, however, regulation by phosphorylation was unanticipated. Understanding of the regulation of cellulose biosynthesis is still vague, but only roles for transcriptional regulation, protein stability, and redox potential have been shown (Doblin et al., 2002). Cellulose synthases, especially those involved in building the primary cell wall, have proven difficult to work with on a biochemical level. Mutational analysis of the phosphorylation sites described in this work may lead to new understanding of the mechanisms regulating these important proteins.

Following the paradigms of receptor Tyr kinases of animals (Schlessinger, 2000), the regulation of plant RLKs by phosphorylation comes as no surprise. Investigation of the catalytic properties of many receptor-like kinases have relied on bacterial expression of fusion proteins and most display strong autophosphorylation activity with varying specificity for Ser or Thr residue. Where these sites were analyzed, 10 or more phosphorylated tryptic peptides were found (Stone et al., 1998; Coello et al., 1999; Lease et al., 2001; Liu et al., 2002). When the mechanism of this autophosphorylation was analyzed, some kinases showed an intermolecular mechanism (Coello et al., 1999; van der Knaap et al., 1999), whereas BRI1, Xa21, and CrRLK phosphorylated mostly or exclusively via an intramolecular mechanism (Schulze-Muth et al., 1996; Oh et al., 2000; Liu et al., 2002). It is worth noting that experiments with kinase domain fusions and full-length proteins can come to remarkably different results. Li et al. (2002) found that *Escherichia coli*–expressed MBP and GST fusions of the BRI1 and BAK1 kinase domains, respectively, have

autophosphorylation activity, whereas the full-length proteins expressed in yeast do not (auto)phosphorylate unless coexpressed (Nam and Li, 2002). Similarly, soluble kinase-dead BRI1 binds soluble BAK1 much less strongly than does wild-type BRI1 (Li et al., 2002), whereas no such effect was seen with the full-length proteins (Nam and Li, 2002) in accordance with the earlier observation of brassinolide binding sites in kinase-dead BRI1 mutant plants (Wang et al., 2001). These differences may reflect the absence of important regulatory factors, either from truncations of the protein or from expression in heterologous systems. Thus, it is extremely important that conclusions made about RLK function are based upon in vivo phosphorylation data.

What does the occurrence of the identified phosphorylation sites tell us about the activation status of the corresponding RLK? For the most common type of regulation, the phosphorylation of residues in the activation loop, it is safe to assume that the modification induces the well-characterized structural rearrangements that render the kinase active (Johnson et al., 1996; Huse and Kuriyan, 2002). Whether this is a regulated step or part of the protein maturation process as for many AGC kinases (Newton, 2003) remains to be established. Several mutations in the activation loops of plant RLKs that lead to aberrant phenotypes have been described (Dievart and Clark, 2003). It has also been shown that the phosphatase KAPP binds to the activation loop in a phosphorylation-dependent manner (Braun et al., 1997). These results underscore the key function of this domain at least in some proteins, but activation mechanisms independent of A loop modification, such as relief of autoinhibition by a domain that folds back onto the kinase core, are very common (Huse and Kuriyan, 2002).

Possible roles for phosphorylation in the highly divergent JM and C-terminal domains are intriguing. For both receptor Tyr and Ser/Thr kinases, a dual function of the JM domain has emerged in the last few years (Hubbard, 2001). The unphosphorylated JM domain contacts the N-terminal lobe of the kinase and keeps it in an inactive conformation, either on its own as for the ephrin receptor EphB2 (Wybenga-Groot et al., 2001) or with the help of an additional protein, such as immunophilin FKBP12 in the case of the TGFβRI (Huse et al., 2001). After kinase activation and phosphorylation of the JM, inhibition of the kinase is released. In addition, these phosphorylated regions then become docking sites for substrate proteins with domains that bind phosphotyrosine (SH2 or PTB domain proteins) or phosphoserine/threonine residues (MH2 or FHA domain proteins). The recent elucidation of the autoinhibited FLT3 kinase structure (Griffith et al., 2004) has added another mechanistically different version of autoinhibition by the JM. It will be interesting to test if any of these mechanisms are realized in plant RLKs. No details are known about intramolecular (autoinhibitory) interactions in RLKs, but two proteins binding to RLKs in a phosphorylation-dependent manner are known: the phosphatase KAPP (Li et al., 1999) and the ubiquitin ligase ARC (Stone et al., 2003). Additional and alternative mechanisms should not be overlooked. For example, the C-terminal hydrophobic motif of AGC kinases folds back onto the kinase domain upon phosphorylation and acts as an allosteric activator (Yang et al., 2002).

We have not been able to find common motifs among the JM and C-terminal phosphorylation sites, and they may indeed be highly individual regulatory sites. Although this is not encouraging in the quest for general paradigms about how plant RLKs work, it may be a requirement for signaling specificity among the hundreds of RLKs in Arabidopsis. Regardless of the current level of understanding, the discovery of phosphorylation in the JM and C-terminal domains of plant RLKs clearly indicates that these regions should be included in experiments designed to elucidate the mechanistic regulation of these proteins.

Few plant proteins in the databases are annotated with bona fide phosphorylation sites. Like in other organisms, the site determination from in vivo phosphorylated protein is very challenging. Many, or perhaps even most, sites are predicted by similarity to related proteins or through computer algorithms and later verified by site-directed mutagenesis. Our experimental data have shown that both strategies are potentially misleading. Although phosphorylation sites are conserved among most putative orthologs and some Arabidopsis paralogs, they do not necessarily stand out to facilitate the choice of sites to mutagenize. Concerning in silico predictions, we have found that approximately one-third of experimentally found sites would have been missed by Netphos (data not shown), one of the best available prediction tools (Blom et al., 1999) (http://www.cbs.dtu.dk/services/NetPhos/). This success rate is significantly lower than that estimated for the prediction of sites in mammalian proteins. In addition, the algorithms used in Netphos have been trained on well-characterized mammalian kinase substrates, and several of the best studied mammalian kinases (PKA and PKC) have no clear homologs in Arabidopsis. Thus, the algorithm probably yields a large number of false positives. To make in silico predictions for plants, we clearly need a much larger number of experimentally verified in vivo phosphorylation sites and (in vitro) determined kinase specificities.

Unlike the open reading frame of a protein, its phosphorylation status is dynamic, and any stocktaking of phosphorylated sites can only be a snapshot during the experimental conditions. Our experimental design, with samples pooled from control and elicitor-treated suspension-cultured cells, should encompass a wide variety of basal, stress-, and elicitor-induced phosphorylation sites. Still, kinases that are only activated by very specific stimuli or only in differentiated tissue would have been inactive and their substrates not found in this data set. The consequences are as follows: (1) none of the cataloged sites can be regarded as the regulatory site of the respective protein, but they may serve as a good first guess for mutagenesis and further biological characterization; (2) this study is only the start of a much larger-scale phosphoproteomic project that takes developmental stages, variously treated plants, differentiated tissues, and other subcellular structures into account. We currently are not able to perform comparisons of phosphoproteomes between two treatment conditions to identify differential components, but this achievement is the goal of our ongoing research.

Criticism has occasionally been raised about the descriptive nature of such large-scale projects, misjudging the complementary nature of hypothesis-driven and data-driven research (Kell and Oliver, 2004). Omic data may not reveal biological mechanisms per se but allow the formulation of new testable hypotheses that are not restrained by established ideas. For example, hypotheses about RLK regulation focused mostly on the

activation loop. With the unanticipated complexity of phosphory-lation sites presented here, new models will have to take other intramolecular interactions into account. The daunting task ahead of us is to understand the function of all these phosphory-lation sites; but on the way to this goal, a careful analysis of the existing data can spark many new ideas.

METHODS

Materials

POROS chromatography materials (Self Pack OligoR3, MC 20, and HQ 20) were purchased from Applied Biosystems (Foster City, CA) and modified porcine trypsin from Promega (Southampton, UK). Microcol-umns were packed in Gelloader tips (Eppendorf, Cambridge, UK) as described by Gobom et al. (1999).

Phosphopeptide Preparation

The isolation of plasma membranes from suspension-cultured *Arabidop-sis thaliana* cells has been described by Nühse et al. (2003a). For this large-scale study, plasma membranes from control cells and cells that have been treated with 100 nM flg22 peptide for 8 min were pooled. Before harvesting of plasma membranes from the last upper phase of the dextran/polyethylene glycol partitioning, 0.01% Brij58 was added to invert the membranes to inside-out orientation. After centrifugation (60 min, 150,000*g*), the membranes were resuspended in 100 mM sodium carbonate, left on ice for 10 min, and then collected by centrifugation (30 min, 100,000*g*). The wash was repeated with 500 mM and then 50 mM ammonium hydrogen carbonate.

The membranes were digested at a 1:50 ratio of trypsin:protein overnight at 37°C. Phosphopeptides were isolated as described by Nühse et al. (2003b). Briefly, the supernatant containing the released peptides was either used directly for IMAC after addition of acetic acid (final concentration 0.2 M) or formic acid was added (final concentration 5%) to the digest and the supernatant purified over an OligoR3 column (i.e., washed with 5% formic acid and eluted with 100 mM acetic acid and 50% acetonitrile).

POROS MC material was prepared according to the manufacturer's instructions and loaded with 100 mM FeCl$_3$ in 100 mM acetic acid. Chromatography was performed as described by Stensballe et al. (2001) with minor modifications. Peptides were batch-bound to the IMAC material by shaking at room temperature in a typical volume of 20 to 50 μL, containing ~2 to 5 μL (settled volume) of POROS MC material. After the incubation, the slurry was packed into Gelloader pipette tips with constricted tips and washed once each with 15 μL of 0.1 M acetic acid and 0.1 M acetic acid/30% acetonitrile, respectively, before elution with dilute ammonia, pH 10.5, or 50 mM ammonium phosphate, pH 9. The majority of phosphopeptides were identified from multiple runs of a two-dimensional liquid chromatographic separation scheme as described by Nühse et al. (2003b). Briefly, 500 μg of plasma membranes were trypsin digested and the released peptides purified over an OligoR3 micro-column as described above but washed with water before elution with 50% acetonitrile. The eluate was diluted with buffer and pH adjusted to final concentrations of 30% acetonitrile/20 mM NH$_4$HCO$_3$, pH ~7. A microcolumn was packed with POROS HQ20 (SAX) and preequilibrated with 30% acetonitrile/25 mM NH$_4$HCO$_3$, pH 7.0 (SAX buffer). The sample was slowly loaded onto the column and the flow-through collected. Twelve fractions were collected by step eluting with 20 μL each of 40 to 500 mM NaCl in SAX buffer. Flow-through and eluate fractions were briefly concentrated in a speed vac to reduce the acetonitrile concentra-tion, brought to 5% formic acid, and desalted on R3 microcolumns. IMAC purification of phosphopeptides was done as described above.

Mass Spectrometry

Automated nanoflow LC-MS/MS analysis was performed using a QTOF Ultima mass spectrometer (Waters/Micromass, Manchester, UK) em-ploying automated data-dependent acquisition. A nanoflow-HPLC sys-tem (Ultimate; Switchos2; Famos; LC Packings, Amsterdam, The Netherlands) was used to deliver a flow rate of 175 nL min^{-1} to the mass spectrometer. Chromatographic separation was accomplished using a 2-cm fused silica precolumn (75-μm i.d. and 360-μm o.d.; Zorbax SB-C18 5 μm; Agilent, Wilmington, DE) connected to an 8-cm analytical column (50-μm i.d. and 360-μm o.d.; Agilent Zorbax SB-C18 3.5 μm). Peptides were eluted by a gradient of 5 to 32% acetonitrile for 35 min.

The mass spectrometer was operated in positive ion mode with a source temperature of 80°C and a countercurrent gas flow rate of 150 L h^{-1}. Data-dependent analysis was employed (three most abundant ions in each cycle): 1 s MS mass-to-charge ratio (*m/z*) 350 to 1500 and max 4 s MS/MS *m/z* 50 to 2000 (continuum mode), 30 s dynamic exclusion.

Raw data were processed using MassLynx 3.5 ProteinLynx (smooth 3/2 Savitzky Golay and center 4 channels/80% centroid) and the resulting MS/MS data set exported in the Micromass pkl format. We performed the peptide identification and assignment of partial post-translational modifications using an in-house version of Mascot version 1.9. All data sets were searched twice, first with relatively large peptide mass tolerances, followed by internal mass recalibration by an in-house software algorithm using theoretical masses from unambiguously iden-tified peptides obtained from the first search. The recalibrated data sets were searched against NCBInr (all species) using the following con-straints: only tryptic peptides with up to three missed cleavage sites were allowed; 0.1 D mass tolerances for MS and MS/MS fragment ions. Phosphorylation (STY), deamidation (NQ), and oxidation (M) were specified as variable modifications. The results were filtered for non-Arabidopsis peptide assignments. External mass calibration using NaI resulted generally in mass errors of <50 ppm, typically 5 to 15 ppm in the *m/z* range 50 to 2000.

Approximately 50 of the MS/MS spectra were inspected for neutral loss, sequence-specific fragmentation patterns, and database match. Most of the spectra showed prominent loss of phosphate. The Mascot search results calculated all hits with a score above 40 to be significant; that is, a nonrandom match at P < 0.05 (the Mascot score is a probabilistic implementation [http://www.matrixscience.com/help/scoring_help.html] of the MOWSE score; Pappin et al., 1993). However, for all manually inspected spectra with a score above 30 and many spectra with a score above 20, the top search hit was in agreement with the spectrum on the basis of characteristic fragmentation. Correct hits may have a low score because of poorer quality spectra. Phosphorylation sites are only marked as unambiguous (bold red type in Supplemental Table 1 online) if sequences with alternative phosphate positions scored at least 10 units lower or were not reported as significant hits. Tryptic peptides that are not unique to one protein are marked in the table. All protein hits were verified in The Arabidopsis Information Resource database (http://www. arabidopsis.org/); no exclusions were made on the basis of predicted function, localization, etc.

Bioinformatics

The following Web-based bioinformatic tools were used: TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) for the prediction of transmem-brane helices; SMART (http://smart.embl-heidelberg.de/) for functional domains; ClustalW (http://www.ebi.ac.uk/clustalw/) for protein sequence alignments; Boxshade (http://www.ch.embnet.org/software/BOX_form.html) for rendering ClustalW results (setting 50% of sequences must agree for shading).

Putative orthologs of phosphoproteins were found by BLAST search of (http://www.ncbi.nlm.nih.gov/BLAST; translated search tblastn) 40 to 60

amino acid sequences surrounding the empirical phosphorylation sites against an EST database (setting est_others; expect 10,000). Where possible, a few amino acids were included in the search that are conserved among several homologs of the phosphoprotein to distinguish unrelated sequences more easily.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under the following accession numbers: AHA2 (At4g30190), MDR1 (At2g36910), MDR4 (At2g47000), MDR8 (At1g02520), MDR17 (At3g62150), AtCesA1 (At4g32410), AtCesA2 (At4g39350), AtCesA3 (At5g05170), AtCesA4 (At5g44030), AtCesA5 (At5g09870), AtCesA6 (At5g64740), AtCesA7 (At5g17420), AtCesA8 (At4g18780), AtCesA9 (At2g21770), and AtCesA10 (At2g25540). For CDPKs, see PlantsP database (http://plantsp.sdsc.edu); for the STP family, see PlantsT (http://plantst.sdsc.edu). CesA3 putative orthologs are as follows: *B. napus* (CD837398); *I. nil* (BJ553825); *M. truncatula* (BQ139762 and BF634138); *G. hirsutum* (AF150630); *P. balsamifera* (BU879114); *T. aestivum* (BQ578769 and BT009438); *Z. mays* (ZmCesA1, AF200525; ZmCesA2, AF200526; ZmCesA4, AF200528; ZmCesA5, AF200529; ZmCesA9, AF200533); *L. japonicus* (AV408250); *O. sativa* (AK072356.1, AK098978, and D48636); *G. max* (BE660209); *L. sativa* (BQ869850); *Rosa* sp (BQ104792); *V. vinifera* (CD005820); *S. propinquum* (BG465505); *S. bicolor* (BG273410); *H. vulgare* (BQ471177); *T. monococcum* (BQ802778). SOS1 (At2g01980); putative orthologs: *P. trem* (CA924192); *G. max* (BE609663); *B. vulgaris* (BQ591881); *S. tuberosum* (BQ510147); *M. chrysanthemum* (BE577626); *O. sativa* (CB676883).

ACKNOWLEDGMENTS

REFERENCES

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796–815.

Bachmann, M., Shiraishi, N., Campbell, W.H., Yoo, B.C., Harmon, A.C., and Huber, S.C. (1996). Identification of Ser-543 as the major regulatory phosphorylation site in spinach leaf nitrate reductase. Plant Cell **8,** 505–517.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. **294,** 1351–1362.

Braun, D.M., Stone, J.M., and Walker, J.C. (1997). Interaction of the maize and Arabidopsis kinase interaction domains with a subset of receptor-like protein kinases: Implications for transmembrane signaling in plants. Plant J. **12,** 83–95.

Cheng, S.H., Willmann, M.R., Chen, H.C., and Sheen, J. (2002). Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family. Plant Physiol. **129,** 469–485.

Coello, P., Sassen, A., Haywood, V., Davis, K.R., and Walker, J.C. (1999). Biochemical characterization and expression of RLK4, a receptor-like kinase from Arabidopsis thaliana. Plant Sci. **142,** 83–91.

Decottignies, A., Owsianik, G., and Ghislain, M. (1999). Casein kinase I-dependent phosphorylation and stability of the yeast multidrug transporter Pdr5p. J. Biol. Chem. **274,** 37139–37146.

Dievart, A., and Clark, S.E. (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. Curr. Opin. Plant Biol. **6,** 507–516.

Doblin, M.S., Kurek, I., Jacob-Wilk, D., and Delmer, D.P. (2002). Cellulose biosynthesis in plants: From genes to rosettes. Plant Cell Physiol. **43,** 1407–1420.

Elortza, F., Nühse, T.S., Foster, L.J., Stensballe, A., Peck, S.C., and Jensen, O.N. (2003). Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. Mol. Cell Proteomics **2,** 1261–1270.

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

Gavin, A.C., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415,** 141–147.

Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., and Roepstorff, P. (1999). Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. J. Mass Spectrom. **34,** 105–116.

Griffith, J., Black, J., Faerman, C., Swenson, L., Wynn, M., Lu, F., Lippke, J., and Saxena, K. (2004). The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. Mol. Cell **13,** 169–178.

Guenther, J.F., Chanmanivone, N., Galetovic, M.P., Wallace, I.S., Cobb, J.A., and Roberts, D.M. (2003). Phosphorylation of soybean nodulin 26 on serine 262 enhances water permeability and is regulated developmentally and by osmotic signals. Plant Cell **15,** 981–991.

Halfter, U., Ishitani, M., and Zhu, J.K. (2000). The Arabidopsis SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3. Proc. Natl. Acad. Sci. USA **97,** 3735–3740.

Ho, Y., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature **415,** 180–183.

Huang, J.Z., Hardin, S.C., and Huber, S.C. (2001). Identification of a novel phosphorylation motif for CDPKs: Phosphorylation of synthetic peptides lacking basic residues at P-3/P-4. Arch. Biochem. Biophys. **393,** 61–66.

Huang, J.Z., and Huber, S.C. (2001). Phosphorylation of synthetic peptides by a CDPK and plant SNF1-related protein kinase. Influence of proline and basic amino acid residues at selected positions. Plant Cell Physiol. **42,** 1079–1087.

Hubbard, S.R. (2001). Theme and variations: Juxtamembrane regulation of receptor protein kinases. Mol. Cell **8,** 481–482.

Huber, S.C., and Huber, J.L. (1996). Role and regulation of sucrose-phosphate synthase in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47,** 431–444.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell **109,** 275–282.

Huse, M., Muir, T.W., Xu, L., Chen, Y.G., Kuriyan, J., and Massague, J. (2001). The TGF beta receptor activation process: An inhibitor- to substrate-binding switch. Mol. Cell **8,** 671–682.

Idriss, H.T., Hannun, Y.A., Boulpaep, E., and Basavappa, S. (2000). Regulation of volume-activated chloride channels by P-glycoprotein: Phosphorylation has the final say! J. Physiol. **524,** 629–636.

Johansson, I., Karlsson, M., Shukla, V.K., Chrispeels, M.J., Larsson, C., and Kjellbom, P. (1998). Water transport activity of the plasma membrane aquaporin PM28A is regulated by phosphorylation. Plant Cell **10,** 451–459.

Johansson, I., Larsson, C., Ek, B., and Kjellbom, P. (1996). The major integral proteins of spinach leaf plasma membranes are putative

aquaporins and are phosphorylated in response to Ca2+ and apo-plastic water potential. Plant Cell **8**, 1181–1191.

Johnson, L.N., Noble, M.E., and Owen, D.J. (1996). Active and inactive protein kinases: Structural basis for regulation. Cell **85**, 149–158.

Kell, D.B., and Oliver, S.G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays **26**, 99–105.

Lease, K.A., Lau, N.Y., Schuster, R.A., Torii, K.U., and Walker, J.C. (2001). Receptor serine/threonine protein kinases in signalling: Analysis of the erecta receptor-like kinase of Arabidopsis thaliana. New Phytol. **151**, 133–143.

Li, J., Smith, G.P., and Walker, J.C. (1999). Kinase interaction domain of kinase-associated protein phosphatase, a phosphoprotein-binding domain. Proc. Natl. Acad. Sci. USA **96**, 7821–7826.

Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E., and Walker, J.C. (2002). BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell **110**, 213–222.

Liu, G.Z., Pi, L.Y., Walker, J.C., Ronald, P.C., and Song, W.Y. (2002). Biochemical characterization of the kinase domain of the rice disease resistance receptor-like kinase XA21. J. Biol. Chem. **277**, 20264–20269.

Maurel, C. (1997). Aquaporins and water permeability of plant membranes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48**, 399–429.

McMichael, R.W., Jr., Kochansky, J., Klein, R.R., and Huber, S.C. (1995). Characterization of the substrate specificity of sucrose-phosphate synthase protein kinase. Arch. Biochem. Biophys. **321**, 71–75.

Nam, K.H., and Li, J. (2002). BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell **110**, 203–212.

Newton, A.C. (2003). Regulation of the ABC kinases by phosphorylation: Protein kinase C as a paradigm. Biochem. J. **370**, 361–371.

Nühse, T.S., Boller, T., and Peck, S.C. (2003a). A plasma membrane syntaxin is phosphorylated in response to the bacterial elicitor flagellin. J. Biol. Chem. **278**, 45248–45254.

Nühse, T.S., Stensballe, A., Jensen, O.N., and Peck, S.C. (2003b). Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Mol. Cell Proteomics **2**, 1234–1243.

Oh, M.H., Ray, W.K., Huber, S.C., Asara, J.M., Gage, D.A., and Clouse, S.D. (2000). Recombinant brassinosteroid insensitive 1 receptor-like kinase autophosphorylates on serine and threonine residues and phosphorylates a conserved peptide motif in vitro. Plant Physiol. **124**, 751–766.

Palmgren, M.G. (2001). Plant plasma membrane H+-ATPases: Power-houses for nutrient uptake. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52**, 817–845.

Pappin, D., Hojrup, P., and Bleasby, A. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. **3**, 327–332.

Peck, S.C., Nühse, T.S., Hess, D., Iglesias, A., Meins, F., and Boller, T. (2001). Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. Plant Cell **13**, 1467–1475.

Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S.P. (2003). A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. **21**, 921–926.

Qiu, Q.S., Guo, Y., Dietrich, M.A., Schumaker, K.S., and Zhu, J.K. (2002). Regulation of SOS1, a plasma membrane Na+/H+ exchanger in Arabidopsis thaliana, by SOS2 and SOS3. Proc. Natl. Acad. Sci. USA **99**, 8436–8441.

Quintero, F.J., Ohta, M., Shi, H., Zhu, J.K., and Pardo, J.M. (2002). Reconstitution in yeast of the Arabidopsis SOS signaling pathway for Na+ homeostasis. Proc. Natl. Acad. Sci. USA **99**, 9061–9066.

Santoni, V., Vinh, J., Pflieger, D., Sommerer, N., and Maurel, C. (2003). A proteomic study reveals novel insights into the diversity of aquaporin forms expressed in the plasma membrane of plant roots. Biochem. J. **373**, 289–296.

Schaller, G.E., and Sussman, M.R. (1988). Phosphorylation of the plasma membrane H+-ATPase of oat roots by a calcium-stimulated protein kinase. Planta **173**, 509–518.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. Cell **103**, 211–225.

Schulze-Muth, P., Irmler, S., Schroder, G., and Schroder, J. (1996). Novel type of receptor-like protein kinase from a higher plant (Catharanthus roseus). cDNA, gene, intramolecular autophosphorylation, and identification of a threonine important for auto- and substrate phosphorylation. J. Biol. Chem. **271**, 26684–26689.

Shah, K., Vervoort, J., and De Vries, S.C. (2001). Role of threonines in the Arabidopsis thaliana somatic embryogenesis receptor kinase 1 activation loop in phosphorylation. J. Biol. Chem. **276**, 41263–41269.

Stensballe, A., Andersen, S., and Jensen, O.N. (2001). Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. Proteomics **1**, 207–222.

Stone, J.M., Trotochaud, A.E., Walker, J.C., and Clark, S.E. (1998). Control of meristem development by CLAVATA1 receptor kinase and kinase-associated protein phosphatase interactions. Plant Physiol. **117**, 1217–1225.

Stone, S.L., Anderson, E.M., Mullen, R.T., and Goring, D.R. (2003). ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen. Plant Cell **15**, 885–898.

Taylor, S.W., Fahy, E., and Ghosh, S.S. (2003). Global organellar proteomics. Trends Biotechnol. **21**, 82–88.

Taylor, W.R. (1997). Residual colours: A proposal for aminochromography. Protein Eng. **10**, 743–746.

van der Knaap, E., Song, W.Y., Ruan, D.L., Sauter, M., Ronald, P.C., and Kende, H. (1999). Expression of a gibberellin-induced leucine-rich repeat receptor-like protein kinase in deepwater rice and its interaction with kinase-associated protein phosphatase. Plant Physiol. **120**, 559–570.

Wang, Z.Y., Seto, H., Fujioka, S., Yoshida, S., and Chory, J. (2001). BRI1 is a critical component of a plasma-membrane receptor for plant steroids. Nature **410**, 380–383.

Washburn, M.P., Wolters, D., and Yates III, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. **19**, 242–247.

Weaver, C.D., and Roberts, D.M. (1992). Determination of the site of phosphorylation of nodulin 26 by the calcium-dependent protein kinase from soybean nodules. Biochemistry **31**, 8954–8959.

Weekes, J., Ball, K.L., Caudwell, F.B., and Hardie, D.G. (1993). Specificity determinants for the AMP-activated protein kinase and its plant homologue analysed using synthetic peptides. FEBS Lett. **334**, 335–339.

Wu, C.C., MacCoss, M.J., Howell, K.E., and Yates, J.R. (2003). A method for the comprehensive proteomic analysis of membrane proteins. Nat. Biotechnol. **21**, 532–538.

Wybenga-Groot, L.E., Baskin, B., Ong, S.H., Tong, J., Pawson, T., and Sicheri, F. (2001). Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. Cell **106**, 745–757.

Yang, J., Cron, P., Thompson, V., Good, V.M., Hess, D., Hemmings, B.A., and Barford, D. (2002). Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. Mol. Cell **9**, 1227–1240.

Phosphoproteomics of the Arabidopsis Plasma Membrane and a New Phosphorylation Site Database

Thomas S. Nühse, Allan Stensballe, Ole N. Jensen and Scott C. Peck

This information is current as of June 18, 2019

Supplemental Data	/content/suppl/2004/09/01/tpc.104.023150.DC1.html
References	This article cites 63 articles, 21 of which can be accessed free at: /content/16/9/2394.full.html#ref-list-1
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm