Quantitative Phosphoproteomic Analysis of Soybean Root Hairs Inoculated with Bradyrhizobium japonicum*

Tran Hong Nha Nguyen‡§¶, Laurent Brechenmacher‡§, Joshua T. Aldrich||, Therese R. Clauss**, Marina A. Gritsenko**, Kim K. Hixson||, Marc Libault‡‡, Kiwamu Tanaka‡§, Feng Yang**, Qiuming Yao§§, Ljiljana Paša-Tolić, Dong Xu§§, Henry T. Nguyen‡§, and Gary Stacey‡§¶¶

Root hairs are single hair-forming cells on roots that function to increase root surface area, enhancing water and nutrient uptake. In leguminous plants, root hairs also play a critical role as the site of infection by symbiotic nitrogen fixing rhizobia, leading to the formation of a novel organ, the nodule. The initial steps in the rhizobia-root hair infection process are known to involve specific receptor kinases and subsequent kinase cascades. Here, we characterize the phosphoproteome of the root hairs and the corresponding stripped roots (i.e. roots from which root hairs were removed) during rhizobial colonization and infection to gain insight into the molecular mechanism of root hair cell biology. We chose soybean (Glycine max L.), one of the most important crop plants in the legume family, for this study because of its larger root size, which permits isolation of sufficient root hair material for phosphoproteomic analysis. Phosphopeptides derived from root hairs and stripped roots, mock inoculated or inoculated with the soybean-specific rhizobium Bradyrhizobium japonicum, were labeled with the isobaric tag eight-plex iTRAQ, enriched using Ni-NTA magnetic beads and subjected to nanoRPLC-MS/MS1 analysis using HCD and decision tree guided CID/ETD strategy. A total of 1625 unique phosphopeptides, spanning 1659 nonredundant phosphorylation sites, were detected from 1126 soybean phosphoproteins. Among them, 273 phosphopeptides corresponding to 240 phosphoproteins were found to be significantly regulated (>1.5-fold abundance change) in response to inoculation with B. japonicum. The data reveal unique features of the soybean root hair phosphoproteome, including root hair and stripped root-specific phosphorylation suggesting a complex network of kinase-substrate and phosphatase-substrate interactions in response to rhizobial inoculation. *Molecular & Cellular Proteomics 11: 10.1074/mcp.M112.018028, 1140–1155, 2012.

Root hairs are known to play an important role in increasing the root surface area for water and nutrient uptake from the soil (1). Found on the surface of the maturation zone of primary and secondary roots, root hairs develop from specialized epidermal cells (trichoblast). New root hair cells continuously develop in the elongation zone, elongating and maturing as the root grows (2). In addition to the critical role in nutrient uptake, the root hair is the primary infection site for symbiotic bacteria (rhizobia) in legume plants. During the first stages of the legume-rhizobium interaction, (iso)flavonoids secreted by the legume induce the rhizobia to synthesize the Nod factor, a specific lipo-chito-oligosaccharide. This bacterial signal molecule elicits a variety of very rapid (within minutes) responses in the root hair cell, including depolarization of the membrane potential and induction of calcium oscillations (3). The root hair then curls to form a shepherd’s crook structure, where the rhizobia become entrapped within the root hair cell wall (4), leading subsequently to an invagination of the root hair plasma membrane and the formation of the tubular infection thread structure by which the bacteria ultimately gain access to the root cortex (5). Root hair physiology under both controlled and biotic/abiotic stress conditions has been studied intensively using a variety of approaches. Numerous root hair mutants have been identified and subsequently linked to the function of various
proteins, such as transcription factors, protein kinases, cell wall-related proteins, etc (6, 7). In soybean, the use of root hair cells as a single cell model for systems biology has been documented (8). A protein reference map for soybean root hair grown under controlled condition was recently established (9). Similarly, a complete transcriptome for this cell type was created in the form of a soybean gene expression atlas that revealed the differential expression of thousands of genes in response to B. japonicum inoculation (10, 11).

Specific protein phosphorylation events are known to be critical for the initiation of the rhizobial infection process. For example, the study of various plant mutants defective in nodulation led to the identification of key protein kinases essential for rhizobial infection and/or nodule formation. These include receptor-like kinases involved in recognition of the Nod factor signal, GmNFR1 (LjnFR1, MtLYK3), and GmNFR5 (LjnFR5, MtNFP) (12, 13). These genes encode LysM-type serine/threonine receptor kinases where the LysM domain is thought to specifically bind the Nod factor, signaling cellular responses. Another important kinase is the nodulation receptor kinase (NORK, MdDM12, SYMRK) (14), a leucine rich repeat receptor-like kinase (LRR RLK), first identified using a non-nodulating alfalfa mutant (15). The induction of calcium oscillations in the root hair cell is an important initial step in Nod factor recognition and infection. These oscillations are likely detected and translated into cellular activity through the action of a calcium calmodulin kinase (CCaMK) (16, 17). A number of other protein kinases are also likely involved in the nodulation process, as evidenced by their differential gene expression following rhizobial inoculation (11).

Given the fact that so many protein kinases are involved in the rhizobial infection process, we hypothesized that there are likely kinase signaling cascades that control aspects of the root hair rapid response to rhizobia. There are several phosphorylation studies of root hair cells. Examples include the study of tyrosine phosphorylation in relation to microtubule dynamics and organization (18) and auxin efflux regulation by the Ser/Thr protein kinase PINOID (19). Other studies focused on phosphorylation events associated with root hair development such as stress-induced mitogen-activated protein kinases (SIMK), which are important for root hair tip growth (20). Another example is cytoplasmic organization and streaming in root hair cells of Limnobium stoloniferum that were shown to be regulated by protein phosphorylation (21). Yet, only a few large-scale phosphoproteomic studies of root hair cells have been reported so far.

The difficulty of isolating root hair cells in sufficient quantity for detailed analysis may explain the lack of large-scale studies of this cell type. In the current study, we report differential protein phosphorylation in soybean root hair cells in response to B. japonicum interaction enabled by the use of ultra-sensitive phosphopeptide LC platform coupled with high-performance MS (22). We employed eight-plex iTRAQ for relative quantification of phosphopeptides in root hairs and stripped roots across nine time points during the first 48 h after rhizobial inoculation. This multiplexed quantitative strategy allowed the analysis of 108 samples in 18 eight-plex iTRAQ experiments, thus dramatically reducing the number of required LC-MS/MS analyses and overall sample consumption. The results suggest a complex pattern of protein phosphorylation events during rhizobial colonization of root hair cells, some of which appear to be specific to this cell type. Moreover, the abundance of 273 phosphopeptides changed significantly upon B. japonicum inoculation, in agreement with the known involvement of phosphorylation cascades in regulating cellular events during rhizobial infection.

**EXPERIMENTAL PROCEDURES**

**Materials**—Trizol Reagent and iTRAQ® Reagent-8Plex Multiplex Kits were obtained from Invitrogen (Carlsbad, CA). Sequencing Grade Modified Trypsin was from Promega (Madison, WI). Coomassie (Bradford) and Bicinchoninic Acid (BCA) Protein Assay were purchased from Pierce Biotechnology. The solid phase extraction column SPE C18, Discovery DSC-18 was from SUPELCO (Bellefonte, PA). Ni-NTA beads were obtained from Qiagen (Valencia, CA). Other chemicals including the Protease Mixture Tablets and Iron (III) chloride were obtained from Sigma-Aldrich (St. Louis, MO).

**Plant Growth, Bacterial Treatment, and Root Hair Isolation**—Plant growth, bacterial treatment, root hair and corresponding stripped root isolation were performed according to (23) (see supplemental Document for more details). Basically, soybean seeds (Glycine max L. (Merrill) cultivar Williams 82) were surface sterilized and sown on agar plates containing 1X B&D nutrients. A cell suspension of B. japonicum strain USDA 110 (OD 600 of 0.8) was used to inoculate the roots of three-day-old soybean seedlings using a mist sprayer. The seedlings were then further incubated for various times (i.e. 0 min, 15 min, 30 min, 1 h, 3 h, 12 h, 24 h, 36 h, and 48 h) following inoculation. Control samples (mock inoculated by spraying with water) were produced at the same time as the treatment samples. After specific incubation times, the whole roots were detached from the shoots and frozen in liquid nitrogen. Root hairs and the corresponding stripped roots were collected for analysis. All samples were kept frozen at −80 °C until protein extraction. Approximately 0.5 to 1.0 g of root hair material was produced from 1000 soybean seedlings. Three independent biological replicates were generated for each time point, resulting in a total of 108 samples for our study (9 time points × 2 conditions × 2 tissues × 3 biological replicates).

**Protein Extraction**—Total protein was extracted using the Trizol Reagent according to the manufacturer’s protocol (Invitrogen) with the addition of the protease and phosphatase inhibitors for each root hair and stripped root sample. We used Protease Mixture Tablets (Sigma S8820), which contain 2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), 14 μM E-64, 130 μM bestatin, 1 μM leupeptin, 0.3 μM aprotinin, and 1 mM EDTA (sodium salt), and phosphatase inhibitors (10 mM β-glycerophosphate, 50 mM NaF, 0.2 mM activated Na3VO4). The resulting protein pellets were stored in 100% ethanol at −80 °C prior to analysis.

**Protein Alklylation and Trypsin Digestion**—Each protein pellet was solubilized in 6 M Guanidine HCl in 50 mM NH4HCO3 pH 7.4, reduced with 10 mM dithiothreitol at 37 °C for 1 h and alkylated with 10 mM iodoacetamide at 37 °C for 1 h in the dark. Protein concentrations were determined using the Bradford Protein Assay according to the manufacturer’s protocol (Pierce Biotechnology). In average, 200–500 μg of total protein were collected for each sample. Each sample was diluted 10-fold with 25 mM NH4HCO3 and doubly digested with trypsin at a ratio of 1:100 (trypsin/substrate) for 6 h at 37 °C (2 h for the experiment).
first digestion and 4 h for the second digestion). The resulting peptides were desalted using SPE C18 (Discovery DSC-18 SUPELCO, Bellefonte, PA); peptide concentrations were determined using the BCA Protein Assay (Pierce Biotechnology).

**Eight-plex iTRAQ Labeling**—For each tissue type (root hair or stripped root), two pooled samples (control and treatment) were generated by combining the digested peptides from nine time points derived from the same condition (control or treatment) for each biological replicate. The pooled samples were included to enable comparison across biological replicates and time points contained in different iTRAQ sets. A 25 µg aliquot of digested peptides for each sample (including the two pooled samples) was subjected to iTRAQ labeling according to the experimental design in Fig. 1 and supplemental Table S1. A single set of eight-plex iTRAQ contained three time points for two conditions (control and treatment) and two pooled samples (i.e., standards), thus yielding a total of nine iTRAQ sample sets for each tissue type. The eight-plex iTRAQ labeling was performed according to the manufacturer’s instructions. Each iTRAQ sample was concentrated and subjected to purification by SPE C18 (Discovery DSC-18 SUPELCO, Bellefonte, PA).

**Phosphopeptide Enrichment**—The eight-plex iTRAQ labeled peptides were subjected to phosphopeptide enrichment using magnetic nickel-nitriloacetic acid (Ni-NTA) beads (Qiagen, Valencia, CA) according to the manufacturer’s protocol with minor modifications for phosphopeptide enrichment (24). The Ni-NTA magnetic beads were washed three times with water and treated with 100 mM EDTA, pH 8.0 for 30 min. After removing the EDTA solution, beads were washed three times with water, and treated with 10 mM FeCl₂ for 30 min. After removing excess metal ions, the beads were washed three times with water and resuspended in 1:1:1 (v/v/v) acetonitrile/methanol/0.01% trifluoroacetic acid. Beads were then conditioned using resuspending/wash buffer containing 80% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid, followed by incubation with peptides (100 µg peptides in 200 µl of 80% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid) for 30 min. The beads were washed three times with the resuspending/wash buffer and the phosphopeptides were eluted with elution buffer containing 50% (v/v) acetonitrile and 2.5% (v/v) ammonia. The eluate containing phosphopeptides was immediately acidified with 10% (v/v) trifluoroacetic acid and concentrated by vacuum centrifugation prior to mass spectrometry analysis.

**nanoRPLC-MS/MS Analysis**—Phosphopeptides from each iTRAQ experiment were subjected to capillary liquid chromatography tandem mass spectrometry using an automated metal-free nanoLC system coupled to a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The automated metal-free nano LC system, which was specifically developed to obtain broad and sensitive analysis of phosphopeptides, is described in detail in (22). MS data were acquired for 200 min, beginning 85 min after sample injection (30 min into gradient). Orbitrap spectra (AGC 5 × 10⁶) were collected from 400–2000 m/z at a resolution of 7500 with subsequent ion trap (AGC 3 × 10⁶) MS/MS acquisition for the 10 most abundant ions. Tandem MS in the ion trap was performed using decision tree guided collisionally induced dissociation (CID; using a collision energy of 35%), or electron transfer dissociation (ETD; using a reaction time of 100 ms). A dynamic exclusion time of 60 s was used to discriminate against previously analyzed ions. Decon2LS (version 1.0.3351.14092, http://omics.pnl.gov/software/DeconTools.php) was used to generate the peak list for iTRAQ searching. Depending on the sample availability, each iTRAQ experiment was analyzed at least twice.

**Phosphopeptide Identification**—The raw HCD, CID and ETD files were searched individually by SEQUEST (version 27) against the combined database derived from the Glycine max ("Glyma1_pep_domains2.fasta," downloaded on December 23, 2008, which includes 77,900 nonredundant predicted peptide sequences) and B. japonicum genome sequences ("Bjaponicum.faa," downloaded on December 2, 2008, which includes 8317 nonredundant peptide sequences) using the forward search mode. Optimized tolerance settings and modifications were as follows: only tryptic peptides were selected (amino- and carboxy-termini were also included), the dynamic modifications of 79.9663 Da (phosphorylation) were applied to serine, threonine, and tyrosine residues; the precursor ion mass tolerance was set at ±0.05 Da; and the fragment ion mass tolerance was set at ±0.5 Da (m/z). A reverse database was searched using all of the data sets from our study and an MSGF spectral probability score of ≤ 1 × 10⁻¹⁰ was found to be an acceptable filter for FDR <1% across all forward searched data sets (25). The location of the phosphorylation site(s) on each phosphopeptide was validated using the A-Score algorithm (26). The final phosphopeptide identifications were constructed with 99% confidence in phosphorylation position derived from the A-Score value [A-Score value >19 or -1 (in case of a single possible site of phosphorylation)]. The spectral information for each phosphopeptide can be found in the Plant Protein Phosphorylation Database at http://p3db.org/createExcelTable2.php (the site works properly under Google Chrome and Firefox only).

**Quantitative Analysis**—MASIC (MS/MS Automated Selected Ion Chromatogram Generator, version v2.5.3923, http://omics.pnl.gov/software/MASIC.php) was used for peak integration of the reporter ions. The reporter ion data containing the intensity of the eight-plex iTRAQ tags were filtered to remove intensity values below 200. Peptides with missed cleavage(s) containing the same phosphorylation site were kept for validation purposes. The log₂ fold-change values (Inoculated/Control) in each time point were calculated for each phosphopeptide. In some cases, the quantitative values of certain time points were not available because of missing phosphopeptide identification in one particular sample or the intensity values failed to pass the cutoff. Only phosphopeptides detected in at least two out of the three biological replicates were used for assessment of significant change in response to B. japonicum. Two independent statistical tests were used in this study. First, the paired t test was employed to identify significant changes between the control and treatment sample among the three biological replicates. In the second test, a t test was performed using the standard deviation of the pooled (standard) sample across the biological replicates and the difference between the control and treatment sample to account for the global sample variability. This dual testing was necessary to prevent measurements with low variance occurring by chance in small sample sizes from being misidentified as significant. The phosphopeptides that passed both t-tests with p value < 0.05 were considered to be significantly regulated. To increase the confidence of the quantitative data, we also included the cutoff for the log₂ fold change values, in which the phosphorylation changes were considered highly significant if the log₂ value ≥0.6 or ≤−0.6 (increasing or decreasing 1.5-fold in phosphorylation activity).

**Protein Annotation**—To identify the corresponding protein for each phosphopeptide, each phosphopeptide was searched against the peptide database derived from the Glycine max and B. japonicum genome sequence. The protein identification challenge, i.e. multiple protein IDs sharing the same peptide, will be discussed in detail in the Results and Discussion section. Phosphoproteins were annotated based on the PFAM, KOG, and PANTHER domain predictions available for the soybean genome sequences ("Glyma1_pep_domains2.fasta"). The annotations were confirmed by comparison to the annotation of the top protein hits from the in-house blast search against the nonredundant protein database (version nr.04, including 11,505,486 sequences) (27). The in-house BLAST...
search (NCBI BLAST version 2.2.18) was performed using default parameters (28). All proteins were manually assigned into 24 different functional categories based on their predicted annotation. A gene enrichment test was performed using a two-side test of the hypergeometric distribution of the target set (phosphoproteins significantly regulated by B. japonicum inoculation) compared with the background set (all phosphoproteins identified in root hairs and stripped roots). The significantly overrepresented functional category was identified with $p < 0.05$.

**Soybean, Arabidopsis and Medicago Phosphoproteome Analysis**—Comparison was made using the phosphoproteome database available in the P3DB database (http://digbio.missouri.edu/p3db/; 29), which includes 13,662 unique phosphopeptides (3856 phosphoproteins) from Arabidopsis thaliana and 4131 unique phosphopeptides (980 phosphoproteins) from Medicago truncatula. We used the whole genome sequences of the Arabidopsis TAIR peptide database (“ArabidopsisPep.fasta,” 35,386 sequences, downloaded April 7, 2011), and the Medicago sequences from the International Medicago Genome Annotation Group and NCBI (“Mit3.0_proteins_20090702_NAME.fasta” and NCBI sequences, which includes 53,543 sequences in total).

**Phosphopeptide Motif Discovery and Protein Characterization**—The phosphorylation site motifs were extracted using the motif-x program (version v1.2, 10.05.06, http://motif-x.med.harvard.edu/motif-x.html) (30) with the default parameters. Signal peptide domains were predicted using SignalP Server (version 3.0, http://www.cbs.dtu.dk/services/SignalP/) (31, 32), and transmembrane domains were predicted using TMHMM Server (version v2.0, http://www.cbs.dtu.dk/services/TMHMM/) (33).

**RESULTS AND DISCUSSION**

**Phosphopeptide Identification**—We profiled the phosphoproteome of the root hairs and the corresponding stripped roots during the first 48 h after inoculation with B. japonicum to gain novel insight into the initial plant cellular response to rhizobial infection. Most functional genomic approaches measure gene/protein/metabolite levels averaged over the whole plant (or multicellular) tissues. Therefore, it is typically impossible to know whether measurements of low gene expression reflect genes that are truly expressed at a low level in all cells or high gene expression, but only in a few cells. To address this issue, we explored the biology of a single, differentiated plant cell type—soybean root hairs. As a result, however, the amount of material available for each time point (iTRAQ channel) was limited to $\sim 25 \mu g$ resulting in a total of $\sim 150 \mu g$ of labeled peptides after pooling the eight samples into a single eight-plex iTRAQ set and desalting (Fig. 1). This is a significant departure from conventional phosphoproteomic studies, which typically involve large amounts of starting material (up to milligrams), thus allowing for extensive peptide fractionation in two-dimensional (2D) LC-MS scheme. In addition, our study includes 108 samples, which would have required hundreds of LC-MS analyses for a label-free quantitative experiment. Hence, our experimental approach was constrained to phosphopeptide enrichment followed by an eight-plex iTRAQ labeling strategy and 1D RPLC-MS/MS analyses (versus 2D approach typically applied to increase overall coverage) because of the limited sample and instrument time availability.

Approximately 32,400 phosphopeptides were identified across all data sets. For each iTRAQ experiment, we combined the CID, ETD, and HCD data to derive a set of unique phosphopeptide identifications (Fig. 2A). In addition to providing quantitative information (via iTRAQ reporter ions), HCD also provided a significant number of unique phosphopeptide identifications (Fig. 2A). Across all 18 iTRAQ experiments, including root hairs and stripped roots, 1625 unique phosphopeptides spanning 1659 nonredundant phosphorylation sites were identified. These phosphopeptides corresponded to 1126 representative phosphoproteins in root hairs and stripped roots. The absence of bacterial proteins is likely due to the low abundance of the bacterial proteins and/or relatively low abundance of (serine, threonine, and tyrosine) phosphoproteins in the cell.

The number of unique phosphopeptides and phosphorylation sites identified in root hairs and stripped roots was particularly impressive considering the limited amount of starting material. Likely, this result is enabled (partly) by our in-house...
developed ultra-sensitive phosphopeptide LC platform. To the best of our knowledge, at present, there is no large-scale phosphoproteomic analysis that has been conducted employing highly specialized plant cell types (e.g. pollen tubes, root hairs, trichomes, and stomatal guard cells, etc.). There are some examples of single cell type analyses published using Arabidopsis cell cultures (34, 35, 36). Thus, our data represent one of the first large-scale phosphoproteomic studies using a single plant-cell type isolated directly from the plant.

Of the 1625 unique phosphopeptides, 95% were singly phosphorylated, 5% were doubly phosphorylated and in only one case, the phosphopeptide had three phosphorylation sites (Fig. 2C). Of the 1659 nonredundant phosphorylation sites, 89.3% were phosphorylated at serine, 10.2% at threonine and 0.5% at tyrosine residue (Fig. 2B). The distribution of phospho-amino acids in our study is consistent with other plant studies, such as those conducted in Arabidopsis and Medicago. However, the abundance of tyrosine phosphorylation in soybean (0.48%) (this study) is slightly lower than in Arabidopsis (4.3%) (37) or Medicago (1.3%) (38). This difference may be attributed to differences in methodology (e.g. phosphopeptide enrichment and/or LC-MS) or biological system, where each cell type, tissue, and organism has a unique phosphoproteome profile.

Impact of Soybean Whole Genome Duplication on Data Analysis—At a certain time point, localization and/or interaction partner, a protein can be phosphorylated at a single or multiple phosphorylation sites. Passing through several steps of protein digestion, phosphopeptide enrichment, fractionation, and mass spectrometry analysis, it is possible that for a specific protein, only a single phosphopeptide is retained and can be detected in the LC-MS analysis. Thus, while at least two peptides are typically required for confident identification of a protein (global proteomics), phosphoproteins are usually inferred directly based on the presence of a single phosphopeptide (phosphoproteomics). Although the SEQUEST search provides the phosphopeptide with its corresponding protein identification, it is possible that more than one protein can share that phosphopeptide. Soybean is a paleotetraploid that is thought to have originated from the combination of two ancient parental genomes, as well as at least one additional whole genome duplication event (39). Consequently, the soybean genome contains many homeologous gene pairs with highly similar sequences (40). Because phosphorylation often occurs in conserved regions of a protein, soybean phosphopeptides are more likely to be shared by homologous protein pairs. Therefore, when identifying the protein corresponding to a certain phosphopeptide, we encountered the issue of a phosphopeptide being shared by multiple proteins.

To identify the corresponding protein for each phosphopeptide, the phosphopeptides in our data set were individually searched against the peptide database derived from the Glycine max genome sequence (supplemental Table S2). It is evident that some phosphopeptides represent mis-cleavaged versions because of limited trypsin digestion (e.g. R.(pS)-TSDAYEK.T and R.(pS)TSDAYEKQTGNRPSSSSPFDV.-). Obviously, short peptides tend to be shared by multiple proteins, whereas the longer peptides are more likely to identify a residue specific to a single protein. As in the example above, R.(pS)TSDAYEK.T is shared by the two proteins: Glyma10g29190 and Glyma20g38100, while R.(pS)TSDAYEKQTGNRPSSSSPFDV.- is unique to Glyma10g29190. Apparently, mis-cleaved peptides can lead to different protein identifications; hence, all mis-cleaved phosphopeptides were included in our phosphopeptide database (supplemental Table S2).

Excluding the cases of the same peptide having different phosphorylation sites, from 1625 unique phosphopeptides, there were 1746 unique peptides for protein identification (The bigger number of unique sequences (peptides) is because of the mis-cleavages). Interestingly, among the 1746 peptides, only 57.4% of the phosphopeptides identified a unique protein; whereas 34.5% of the phosphopeptides were shared by two proteins and 8.1% were shared by more than two proteins (Fig. 3). In one case, a single phosphopeptide with 10
residues was found to be shared by 21 different proteins (supplemental Table S2). Given that more than 40% of the phosphopeptides were shared by two or more proteins, it is difficult to make the connection between the phosphoproteomic results and specific elements of root hair biology. To overcome this obstacle, we investigated the relationship between the proteins sharing the same phosphopeptides based on multiple protein sequence alignment and annotation analysis. The sequences of proteins sharing the same phosphopeptides were aligned by the blast2 program (BLAST 2.2.18), and their biological function (annotation) was predicted based on the PFAM, KOG and PANTHER domain predictions and blast search against the nonredundant protein database (nr database).

The majority of proteins in our data set were annotated by domain predictions and blast searches. In most cases, proteins sharing the same phosphopeptides have the same predicted function. Many have high sequence similarity (>75%) (supplemental Fig. S1) suggesting these might be homologous gene pairs. Within proteins sharing the same phosphopeptide, 127 cases had no annotation information (“Unknown”). By sequence alignment, most of these are likely paralogs based on their high sequence similarity (supplemental Fig. S1). Taken together, the majority of proteins sharing the same phosphopeptides had the same biological functions and likely the same domain structures. Therefore, to derive biological information from our data set, for each phosphopeptide, only a single representative protein from each protein group sharing that peptide was used for functional analysis.

Shared Peptide Aspect of Soybean (Glycine max), Arabidopsis thaliana and Medicago truncatula Phosphoproteome—To investigate the issue of unique and shared phosphopeptides in other plant phosphoproteomes, we performed the same analysis using published data from Arabidopsis and Medicago. Note, however, that these data were derived from tissues other than root hair cells. The P3DB database (29) contains a total of 6151 unique Arabidopsis peptides, originating from 13,662 unique phosphopeptides derived from 14 studies, while a total of 2202 unique peptides from 4131 unique phosphopeptides were identified from a single study in Medicago (38). These data were extracted and searched against either the Arabidopsis TAIR peptide database or the Medicago sequence available through NCBI and the International Medicago Genome Annotation Group.

Interestingly, compared with 57.4% in soybean, 95.7% of the Arabidopsis and 83.3% of the Medicago phosphopeptides were uniquely associated with specific proteins. Although 34.5% of soybean phosphopeptides were shared by two proteins, only 14.2% were shared in Medicago, and just 2.9% in Arabidopsis (Fig. 3). These results likely reflect the relative genome size (and coding capacity) of these three genomes [e.g. soybean is predicted to have roughly $3\times$ more ORFs than Arabidopsis (40)] and, most importantly, the recently duplicated nature of the soybean genome where paralogs show a high level of sequence identity. In addition, the Arabidopsis phosphoproteome has been characterized in more depth because of the higher number of studies and better annotation of this genome. Although trypsin was the only digestion enzyme used in this study, data from Medicago were generated using the combination of variety of digestion enzymes i.e. trypsin, ArgC, AspN, GluC, and LysC (38). By combining different digestion enzymes, it is likely that a variety of peptide populations will be generated, increasing the likelihood that a peptide is unique for a specific protein.

The function of 1126 plant proteins corresponding to 1625 unique phosphopeptides was searched based on PFAM, KOG and PANTHER domain predictions available for soybean. The annotation was confirmed by blast searching against the nonredundant protein database (version nr.04). These proteins were then classified into 24 different categories based on their predicted functions (Fig. 4). The largest functional groups were DNA/RNA-related proteins (26.7% by combining six categories together), signal transduction (13.3%), miscellaneous group (proteins with multiple functions, 9.9%), and protein trafficking (5.4%). No annotation could be assigned to 20.8% of the phosphoproteins identified.

Root Hair-specific Phosphorylation and Stripped Root-specific Phosphorylation—Proteins can serve many different functions depending on their cellular localization, temporal expression, interaction partner(s), regulatory mechanisms, etc. Obviously, protein phosphorylation can contribute to the tight regulation of protein networks, leading to a specific physiology and function of different cell types and tissues. In this study, the phosphoproteomes of root hair cells and the
corresponding stripped roots were qualitatively compared. With nine iTRAQ experiments covering nine time points and three biological replicates for root hair or stripped root, a phosphoprotein (or phosphorylation site) was considered to be specific to root hair or stripped root if it was detected in at least three out of nine iTRAQ experiments in root hair but was not detected at any time in stripped root, and vice versa. Using this cutoff, 86 and 60 phosphoproteins were specifically found only in root hairs and stripped roots, respectively (Fig. 5). Among these, several protein transporters were shown to be specific to root hairs, consistent with the key role of these cells in water and nutrient uptake. In addition, there were many proteins involved in signal transduction, cytoskeleton, protein trafficking, stress-related proteins, etc. that were specifically detected in root hair.

Interestingly, there were 47 phosphoproteins with multiple phosphorylation sites detected in both root hairs and stripped roots. Among these, 24 and 23 proteins with one or more phosphorylation sites appeared to specifically occur in root hairs or stripped roots, respectively (Table I). In this list, six calmodulin binding proteins were detected with multiple phosphorylation sites; one of the sites was specific for root hairs (five proteins) or stripped roots (two proteins). This list of proteins also included Ser/Arg-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins, which are part of the plant mRNA splicing machinery, a major target of regulatory phosphorylation (36). These data raise the possibility that root hair/stripped root-specific choice of phosphorylation sites could be important in the regulation of protein function.

These tissue-specific phosphoprotein patterns could either be because of very specific expression/modification of these proteins in root hairs or to the dilution effects that are inherent in analyzing the protein content of whole roots. Because of the complexity of cell types in stripped roots, which include epidermis, pericycle, endodermis, etc, the data have to be interpreted with caution. Indeed, the ability to specifically enrich for proteins only expressed in a single cell type and in response to *B. japonicum* inoculation is a key reason for our choice of the root hairs for phosphoproteome analysis.

**Comparison of the Root Hair Phosphoproteome and Proteome**—The first proteomic analysis on soybean root hairs was performed by Wan *et al.*, 2005 (23) using 2D PAGE coupled with MS analysis. Among 57 proteins identified in Wan *et al.* study, 10 proteins were detected in this phosphoproteomic study, including a stress-induced gene H4, phosphoenolpyruvate (PEP) carboxylase, S-adenosylmethionine synthetase 3, enolase, late embryogenesis abundant (LEA) protein, cyclophilin, phosphoglucomutase, putative heat shock 70 kDa protein mitochondrial precursor, 2,3-bisphosphoglycerate independent phosphoglycerate mutase, and cinnamyl-alcohol dehydrogenase. Interestingly, a phosphopeptide corresponding to S-adenosylmethionine synthetase 3, which was found to be more abundant in root hairs compared with stripped roots (23), was found specifically in the
root hair samples analyzed. This result suggests a specific role for this protein in the root hair cell.

Phosphorylation Site Motifs in Comparison to Other Plant Phosphoproteomes—The phosphorylation motifs of 1659 non-redundant soybean phosphorylation sites characterized in our study were extracted using the motif-x program (30). In order to compare the potential consensus sequences among different plants, all phosphopeptides from our soybean study, as well as those from previously published studies with Arabidopsis and Medicago (downloaded from the P3DB database), were used for motif extraction against the background database generated by combining 9000 random protein sequences derived from the annotation of these three genomes. All the phosphorylation motifs identified from Arabidopsis, Medicago and soybean can be found in supplemental Table S3D. Among 44 phosphorylation motifs identified in our soybean data, we found 17 common phosphorylation motifs shared by the three plant species. Interestingly, in addition to the ten distinct motifs unique to soybean, legume plants share nine distinct motifs that were not found in Arabidopsis (Table II). These new motifs might indicate unique features of the legume root system, in particular related to symbiosis.

Eight-plex iTRAQ Quantitative Analysis and the Protein Phosphorylation Changes in Response to Bradyrhizobium japonicum Inoculation—We did not identify any phosphopeptides uniquely associated with only inoculated or mock-inoculated (control) samples. Among the 1860 phosphopeptides (including miscleaved peptides) in our data set, 1348 and 1284 phosphopeptides from nine iTRAQ experiments of root hairs and stripped roots, respectively, had reporter ions that were quantifiable (supplemental Table S5A), i.e. having the reporter ion information for control and treatment sample across three biological replicates at a given time point. The fold change for each phosphopeptide at each time point was log₂ transformed prior to analysis. Overall, the data showed higher consistency between biological replicates in the early time points comparing to the later time points (supplemental Fig. S2 and S3). This can be explained by the nature of our experimental design, in which the gap between the early time points was significantly shorter than at the later time points. Hence, the differential phosphorylation patterns associated with inoculation can be better captured for the early time points (36).

The abundance of 299 nonredundant phosphopeptides, corresponding to 272 phosphoproteins, were found to change significantly at least at one time point in root hairs or stripped roots (supplemental Table S5B) using dual t test on all time points. Applying a stringent filter, log₂[fold change]>0.6 or <=-0.6, 273 phosphopeptides, corresponding to 240 phosphoproteins, were found to be significantly regulated (supplemental Table S5C). Among these, the abundance of 37 phosphoproteins were significantly changed in both tissues, whereas the abundance of 124 and 79 phosphoproteins varied significantly only in root hairs and stripped roots, respectively (Fig. 6).

It should be noted that we cannot discern whether the quantitative changes in the abundance of these 240 phosphoproteins is because of a change in the relative abundance of the protein or because of changes in the activities of specific kinases or phosphatases. Of course, from a biological perspective, both types of changes are of interest. However, this issue is one of the intrinsic problems in MS-based large-scale quantitative phosphoproteomic studies (41). At the peptide level, the phosphorylated and nonphosphorylated versions of a peptide are likely missing for the comparison (because of phosphopeptide enrichment methods, nature of the protein abundance, peptide separation and mass spectrometry analysis approaches). At the protein level, because of the duplication of the soybean genome, about 70% of the phosphopeptides significantly regulated by B. japonicum in our studies...
are shared by two or more proteins (supplemental Table S5B), further complicating any attempt to relate the abundance of a specific phosphopeptide to the abundance of its corresponding protein(s). In one specific case, we were able to identify a phosphopeptide that is unique to one protein (phosphoglucomutase Glyma05g34790.1) and for which we have quantitative protein data from the study of Wan et al., 2005 (GenBank ID Q9SM60, Table II) (23). In our study, the phosphopeptide

| No | Representative phosphoprotein | Length | Common PhosSite(s) | PhosSite(s) specific (a:b) |
|----|--------------------------------|--------|--------------------|---------------------------|
| 1  | Glyma13g04480.1 Glycosyl transferase family | 661    | S561               | S586 (3:0)               |
| 2  | Glyma02g1220.1 Linker histone H1 and H5 family | 177    | S104*; S176        | S144 (4:0)               |
| 3  | Glyma11g36120.1 Protein related to chromatin condensation | 683    | S319; S598; S631*  | S57 (3:0)                |
| 4  | Glyma01g02810.1 Myosin 1 | 978    | S7*; S807          | S931* (7:0)              |
| 5  | Glyma02g15590.1 Calmodulin binding protein | 536    | S486; S96; S96S101; S47* | S351 (4:0)              |
| 6  | Glyma03g33560.1 Calmodulin binding protein | 478    | S305*; S312; S314*; S447* | S378 (3:0)              |
| 7  | Glyma04g01820.1 Calmodulin binding protein | 1063   | S910; S599*; S997  | T900 (6:0); T914 (3:0)  |
| 8  | Glyma09g40740.1 Calmodulin binding protein | 184    | S16*               | S23* (3:0)               |
| 9  | Glyma10g07410.1 Late embryogenesis abundant protein | 450    | S117               | S142 (4:0)               |
| 10 | Glyma13g36780.1 Seed biotin-containing protein | 644    | S623               | S530* (5:0); T631 (6:0) |
| 11 | Glyma13g36780.1 Seed biotin-containing protein | 644    | S623               | S530* (5:0); T631 (6:0) |

Phosphosite specific for root hair

| No | Representative phosphoprotein | Length | Common PhosSite(s) | PhosSite(s) specific (a:b) |
|----|--------------------------------|--------|--------------------|---------------------------|
| 1  | Glyma09g02160.1 Endo-1,4-beta-glucanase | 619    | T16                | S18 (0:3); S37 (0:5)      |
| 2  | Glyma02g45660.1 Microtubule association protein (MAP) | 583    | S533               | S511 (0:4)               |
| 3  | Glyma01g04650.1 Microtubule association protein (MAP) | 197    | S71                | S75 (0:6)                |
| 4  | Glyma11g34830.1 Regulatory protein MLP/LIM proteins | 201    | S93; S95*          | S178 (0:3)               |
| 5  | Glyma04g02830.1 Calmodulin binding protein | 837    | S658; S778         | S736 (0:3)               |
| 6  | Glyma05g33880.1 Calmodulin binding protein | 217    | S202               | S186 (0:3)               |
| 7  | Glyma08g42920.1 Nitrate-induced NOI protein | 105    | T24                | S42 (0:4)                |
| 8  | Glyma13g43360.1 Harpin-induced protein 1 (Hin1) | 244    | S18                | S77 (0:4)                |
| 9  | Glyma10g35520.1 Aquaporin | 297    | S289S292*; S292*   | S289 (0:3)               |
| 10 | Glyma08g43740.1 Heterogeneous nuclear ribonucleoprotein | 480    | S233*; S310; S421  | S281 (0:3)               |
| 11 | Glyma09g02160.1 Endo-1,4-beta-glucanase | 619    | T16                | S18 (0:3); S37 (0:5)      |
| 12 | Glyma02g45660.1 Microtubule association protein (MAP) | 583    | S533               | S511 (0:4)               |
| 13 | Glyma01g04650.1 Microtubule association protein (MAP) | 197    | S71                | S75 (0:6)                |
| 14 | Glyma11g34830.1 Regulatory protein MLP/LIM proteins | 201    | S93; S95*          | S178 (0:3)               |
| 15 | Glyma04g02830.1 Calmodulin binding protein | 837    | S658; S778         | S736 (0:3)               |
| 16 | Glyma05g33880.1 Calmodulin binding protein | 217    | S202               | S186 (0:3)               |
| 17 | Glyma08g42920.1 Nitrate-induced NOI protein | 105    | T24                | S42 (0:4)                |
| 18 | Glyma13g43360.1 Harpin-induced protein 1 (Hin1) | 244    | S18                | S77 (0:4)                |
| 19 | Glyma10g35520.1 Aquaporin | 297    | S289S292*; S292*   | S289 (0:3)               |
| 20 | Glyma08g43740.1 Heterogeneous nuclear ribonucleoprotein | 480    | S233*; S310; S421  | S281 (0:3)               |

Phosphosite specific for stripped root

| No | Representative phosphoprotein | Length | Common PhosSite(s) | PhosSite(s) specific (a:b) |
|----|--------------------------------|--------|--------------------|---------------------------|
| 1  | Glyma09g02160.1 Endo-1,4-beta-glucanase | 619    | T16                | S18 (0:3); S37 (0:5)      |
| 2  | Glyma02g45660.1 Microtubule association protein (MAP) | 583    | S533               | S511 (0:4)               |
| 3  | Glyma01g04650.1 Microtubule association protein (MAP) | 197    | S71                | S75 (0:6)                |
| 4  | Glyma11g34830.1 Regulatory protein MLP/LIM proteins | 201    | S93; S95*          | S178 (0:3)               |
| 5  | Glyma04g02830.1 Calmodulin binding protein | 837    | S658; S778         | S736 (0:3)               |
| 6  | Glyma05g33880.1 Calmodulin binding protein | 217    | S202               | S186 (0:3)               |
| 7  | Glyma08g42920.1 Nitrate-induced NOI protein | 105    | T24                | S42 (0:4)                |
| 8  | Glyma13g43360.1 Harpin-induced protein 1 (Hin1) | 244    | S18                | S77 (0:4)                |
| 9  | Glyma10g35520.1 Aquaporin | 297    | S289S292*; S292*   | S289 (0:3)               |
| 10 | Glyma08g43740.1 Heterogeneous nuclear ribonucleoprotein | 480    | S233*; S310; S421  | S281 (0:3)               |

Phosphorylation sites that were significantly regulated upon B. japonicum inoculation.

(a:b): a and b represents the number of time the phosphorylation site was detected in the iTRAQ experiments of root hairs and stripped roots, respectively.

TABLE I

List of proteins showing specific phosphorylation sites differentially modified in either root hair cells or stripped roots

- Phosphorylation sites that were significantly regulated upon *B. japonicum* inoculation.
- (a:b): a and b represents the number of time the phosphorylation site was detected in the iTRAQ experiments of root hairs and stripped roots, respectively.
Table II

Soybean phosphorylation motifs detected in our study

All phosphorylation motifs identified in our study and their overlap with those present in Arabidopsis and Medicago phosphopeptides identified from previous studies. The localized phosphorylated residue is indicated in bold underline letter (S or T), and . indicates any amino acid.

| No | Soybean motifs | Found in Medicago | Found in Arabidopsis |
|----|----------------|-------------------|----------------------|
| 1  | .R. .SP. .     | x                 | x                    |
| 2  | .RS. .S. .E.  | x                 | x                    |
| 3  | .SP. .R. .     | x                 | x                    |
| 4  | .R. .SP. .     | x                 | x                    |
| 5  | .L. .S. .     | x                 | x                    |
| 6  | .RS. .S. .     | x                 | x                    |
| 7  | .R. .S. .S. .S.| x                 | x                    |
| 8  | .G. .S. .     | x                 | x                    |
| 9  | .R. .S. .S. .S.| x                 | x                    |
| 10 | .RS. .S. .     | x                 | x                    |
| 11 | .S. .R. .S.   | x                 | x                    |
| 12 | .KS. .S. .     | x                 | x                    |
| 13 | .S. .R. .     | x                 | x                    |
| 14 | .SP. .R. .     | x                 | x                    |
| 15 | .S. .PK. .     | x                 | x                    |
| 16 | .SPR. .S. .   | x                 | x                    |
| 17 | .R. .S. .     | x                 | x                    |
| 18 | .SP. .S. .     | x                 | x                    |
| 19 | .SPS. .S.     | x                 | x                    |
| 20 | .SP. .S. .     | x                 | x                    |
| 21 | .R. .S. .D. . | x                 | x                    |
| 22 | .R. .S. .S.   | x                 | x                    |
| 23 | P. .S. .S. .S.| x                 | x                    |
| 24 | S. .S. .S. .S.| x                 | x                    |
| 25 | .S. .D. .D.   | x                 | x                    |
| 26 | .RS. .S. .     | x                 | x                    |
| 27 | .RS. .S. .S.  | x                 | x                    |
| 28 | .SPR. .S. .   | x                 | x                    |
| 29 | .SPR. .S. .   | x                 | x                    |
| 30 | .SP. .S. .SP. | x                 | x                    |
| 31 | .SP. .S. .SP. | x                 | x                    |
| 32 | KS. .S. .     | x                 | x                    |
| 33 | R. .S. .R. .S.| x                 | x                    |
| 34 | SP. .S. .     | x                 | x                    |
| 35 | SP. .S. .SP.  | x                 | x                    |
| 36 | .S. .D. .E. . | x                 | x                    |
| 37 | .R. .S. .     | x                 | x                    |
| 38 | .S. .S. .     | x                 | x                    |
| 39 | .S. .SP. .SP. | x                 | x                    |
| 40 | .SP. .S. .SP. | x                 | x                    |
| 41 | .SP. .S. .SP. | x                 | x                    |
| 42 | TP. .S. .     | x                 | x                    |
| 43 | .P. TP. .TP.  | x                 | x                    |
| 44 | PT. . .PT. .PT.| x                 | x                    |

derived from this protein was more abundant 1 h after inoculation but was significantly less abundant 12 h after inoculation. In contrast, the result of Wan et al. (23) showed a significant increase in the abundance of this protein only 6 h after inoculation, with the protein abundance unchanged at the other time points tested (i.e. 0, 3 and 12 h). Therefore, for this specific case, the relative abundance of the phosphopeptide does not seem to be because of changes in overall protein abundance. In a previous phosphoproteomic study done by Smith et al. (42), they addressed the issue of relating phosphopeptide abundance to protein abundance by showing that there were a number of proteins that have multiple phosphopeptides, only some of which changed significantly in the treatment. One would assume that if the changes in phosphopeptides were solely because of changes in protein abundance, then all phosphopeptides derived from a single protein would change as the protein level changed. Similar to the logic of Smith et al. (42), as shown in supplemental Table S5A, we found many cases where the levels of a single phosphopeptide changed in abundance, whereas other sites from the same proteins showed no significant change. Clearly, for those proteins of interest to address a specific hypothesis, more work would be required to address these relationships.

Differential Phosphorylation Patterns—The majority of differential phosphopeptides increased in abundance in both root hairs and stripped roots very early, at 30 min to 1 h after B. japonicum inoculation (Fig. 6). At the later time points (12 h to 48 h after B. japonicum inoculation), we typically saw the reduction in phosphorylation, and only a few phosphopeptides showed increased abundance at these later time points (Fig. 6). Noticeably, most of phosphorylation changes occurred rapidly, i.e. within an hour after inoculation, which is consistent with the initiation of signaling between the host and symbiont. Differential phosphorylation detected for T0 time point can be explained by the experimental design that included bacterial and mock inoculation. In this case, the T0 time point represents the fastest time needed for (mock) inoculation of the plants and their transfer to liquid nitrogen, which was estimated to be 1–5 min, depending on how many seedlings were harvested. Hence, it is plausible that the time required for treatment would be sufficient to induce some changes in the soybean roots.

The 240 phosphoproteins whose phosphorylation level changed significantly in response to B. japonicum inoculation were classified into 24 functional categories based on their predicted biological function (Fig. 4). The largest functional groups were DNA-RNA related protein (27.6% including transcription factors, RNA processing/modification, translation, chromatin-related proteins, DNA binding proteins), signal transduction (17.9%), unknown (15.8%), protein trafficking (8%), and cytoskeleton (5%). The specific annotation of each phosphoprotein is included in supplemental Table S5C. A visual display of these predicted functions in the context of cellular metabolism is shown in supplemental Fig. S4 using Mapman (43). We performed a gene enrichment test to identify significantly overrepresented functional categories among the group of B. japonicum-responsive phosphoproteins, as compared against the background data set containing all the phosphoproteins identified in root hairs and stripped roots. Three functional categories including signal transduction, pro-
tein trafficking and translation were found to be significantly overrepresented (Fig. 4). These results suggest that these three functional categories are the most important groups that change in response to \textit{B. japonicum} inoculation.

**Phosphorylation Activity of Nodulation Related Proteins**—A few soybean protein kinases have been experimentally shown to play critical roles in Nod factor recognition and subsequent cellular signaling events [e.g. \textit{GmNFR1, GmNFR5} (44)]. Similar important roles can be deduced for other soybean kinases by their orthology to proteins studied in other legumes, including soybean orthologs of NORK (\textit{GmNORK}, Glyma09g33510 and Glyma01g02450) and DMI3 (\textit{GmDMI3}, Glyma15g35070, and Glyma08g24360) (40). Likely because of their low abundance, none of these kinases were detected in our analysis. A nodulin-like protein (Glyma16g06020), which encodes a monocarboxylase transporter, showed differential phosphorylation in root hairs at 1 h and 24 h after rhizobium infection. We also investigated the phosphorylation activity of more than 100 soybean proteins predicted to play a role in nodulation (10, 40). However, none of these were detected in our quantitative data. This is not entirely surprising, given that these proteins may not be regulated by phosphorylation and/or are present at low levels.

**DNA/RNA-related Proteins**—Many phosphoproteins involved in a variety of DNA/RNA process were found to be differentially phosphorylated including 15 ribosomal proteins, five bZIP transcription factors, 13 serine/arginine rich splicing factors, etc. These results suggest that transcription and translation are major targets for regulatory phosphorylation early in the rhizobial-root hair interaction. This would certainly be consistent with the large cellular changes associated with the root hair infection process.

**Rhizobium-induced Differential Phosphorylation of Proteins Involved in Signaling**—Signal transduction was the largest group of proteins whose phosphorylation level was affected by \textit{B. japonicum} inoculation (supplemental Fig. S5). We found 22 protein kinases, four protein phosphatases and one protein phosphatase inhibitor, nine G protein-related proteins, and four proteins involved in hormone signaling within this group (supplemental Fig. S5). These represent candidate gene targets to study in detail with regard to their role in the nodulation process.

**Evidence of a Putative Kinase Cascade within 1 h of \textit{B. japonicum} Inoculation**—Three protein kinases (a putative leucine-rich repeat receptor like protein kinase Glyma01g43340.1, a serine/threonine protein kinase Glyma10g31630 and a casein kinase Glyma10g32490) showed increased phosphorylation in both root hairs and stripped roots after inoculation. In root hairs, the phosphorylation status of 16 different protein kinases and four protein phosphatases was significantly changed within 1 h after \textit{B. japonicum} inoculation (Fig. 7A). We suggest that these changes might reflect the coordinated activity of one or more kinase cascades involved in the infection process. How these proteins fit into the context of root hair signaling, and their relationships with other well-known receptor kinases and protein kinases, would be an attractive topic for further study.

**The Contribution of Calcium Signaling to Soybean–\textit{B. japonicum} Interaction**—Calcium signaling is essential for root hair deformation and infection thread development; many proteins are thought to be involved in generating and interpreting the \textit{Ca}^{2+} signature (45, 46). Among 22 protein kinases in our quantitative phosphoproteomic data, the phosphorylation of six \textit{Ca}^{2+}/calmodulin-dependent protein kinases were increased within one hour after \textit{B. japonicum} inoculation. Moreover, five calmodulin binding proteins, which are classified within the miscellaneous group (Fig. 4) because of their multiple functions, showed similar increases in phosphorylation. The phosphorylation status change of these calcium-related proteins is consistent with the known critical role of calcium signaling during the rhizobium infection process.

**The Contribution of Hormones to the Plant Root Responses**—The phosphorylation level of four proteins was significantly

![Fig. 6](image-url)
changed by inoculation but only in the stripped root tissue. A NINJA-family protein (Glyma01g34820), a negative regulator of the jasmonic acid response (47) showed a decrease in phosphorylation 48 h after inoculation. Jasmonic acid functions in both plant growth/development and defense (48). Hence, one can postulate either a role for phosphorylation in controlling nodule development or in the possible decrease in the plant defense response to allow for rhizobial infection.

Three auxin-related phosphoproteins were also significantly regulated in stripped roots. The first two are encoded by splicing variants of a dormancy/auxin associated protein (Glyma10g29190). The third is an auxin efflux carrier component (Glyma07g22340). Many components of the auxin efflux carrier system are activated by an increase in phosphorylation (49); thus, the increase in phosphorylation activity of Glyma07g22340 is likely related to the activation of auxin transport. Given that cell division in the root cortex is dependent on the regulation of auxin transport (50), these phosphoproteins are likely involved in the signaling pathway for nodule primordium development initially stimulated by the rhizobial Nod factor signal.

**Phosphoproteins Involved in the Root Hair Deformation and Infection Thread Progression**—A root hair cell extends by rapid tip elongation, called tip growth or polar growth (51, 52), in a way that is distinctive from the normal radial growth exhibited by most other plant cells. In response to rhizobium infection, the root hair undergoes deformation and curling, extending in a polar fashion toward the Nod factor application point (53, 54). In both cases, the root hair expands its volume by extension of the plasma membrane and cell wall at the growing tip of the cell (55, 56). This growth process is complex, requiring the coordination of cellular signaling events, membrane trafficking via exocytosis and endocytosis, cytoskeleton reorganization and regulation, and an increase in turgor pressure and water uptake (via aquaporin channels and other transporters). Interestingly, many key

![Fig. 7. Some classes of soybean phosphoproteins in root hair that have the significant change in phosphorylation level in response to *B. japonicum* inoculation. The pattern of the changes in phosphorylation level (fold change, y axis) was investigated in four groups of proteins (A, protein kinases; B, G-related proteins; C, cytoskeleton components; D, ion transporters) across a time course of 0 to 48 HAI (x axis). The standard errors computed from three independent biological replicates were shown in the graph. Details are available in supplemental Table S5.](https://example.com/fig7.png)
components involved in this complex polar root hair cell growth and deformation were identified in our study as elaborated below.

Cell Membrane-associated Signaling Apparatus and Cytoskeleton Regulation—Among 39 signaling phosphoproteins, nine G protein/GTPase related phosphoproteins were found; seven of which were specifically phosphorylated in response to *B. japonicum* at the early time points in root hair cells (Fig. 7B); whereas two others showed decreased phosphorylation in stripped roots. Several are the Rac/Rho-like small GTPases (ROPs) that may regulate actin dynamics and calcium gradient formation (57, 58). In addition, eight phosphoproteins involved in cytoskeleton reorganization and regulation were identified in root hairs, including β-tubulin (component of microtubules), regulatory protein MLP, actin regulatory proteins (villin/gelsolin), and myosin related proteins (Fig. 7C). It is noteworthy that actin regulation, which is required for root hair deformation (59) and infection thread development (21), is controlled by the activation of the ROPs (60, 61). Hence, the presence of differential phosphorylation of these proteins may be essential for the rapid responses of root hair cells to *B. japonicum* inoculation.

Membrane trafficking plays a crucial role in root hair tip growth by providing lipid, protein and cell wall components to the newly developing plasma membrane and cell wall (through exocytosis) and by recycling membrane components (through endocytosis). Among 20 trafficking-related proteins, the phosphorylation level of 16 were significantly changed in root hair cells upon *B. japonicum* inoculation (supplemental Fig. S5), including four ADP-ribosylation factor (ARF) GTPase activators, proteins that regulate vesicular trafficking and also actin remodeling (62, 63). We also found several key phosphoproteins involved in vesicle fusion, including annexin, syntaxin, and vesicle coat complex AP-1. Other trafficking proteins, such as EPSIN/ENT-related protein, clathrin, and Sec14, which regulates lipid metabolism and vesicle transport (64), were also detected.

**Turgor Pressure and Water-potential Regulation**—The phosphorylation level of many transporters involved in cell ion/water-potential regulation was significantly changed upon *B. japonicum* inoculation in root hair cells, but not in stripped roots. These specific phosphoproteins included five transporters predicted to function in ion metabolism/transport (two sodium/proton antiporter, a cation transport ATPase, a potassium transporter, and a voltage-dependent anion-selective channel) (Fig. 7D), as well as other transporters (i.e. aquaporin and ABC transporter). Regulation by phosphorylation of ABC transporters is important for regulating both the activity and stability of such transporters (65). In yeast, at least four phosphorylation sites exist in the ABC core domain involved in positive phosphorylation regulation; and one phosphorylation site (Ser251) is responsible for negative regulation of the yeast vacuolar membrane protein Ycf1p (66). Regulation of transporters by phosphorylation appears to play a role in quickly and reversibly activating/inactivating the transport channels, leading to the rapid control of water potential and turgor pressure, which is necessary for root hair tip growth (56, 67, 68). Hence, phosphorylation regulation of transporters likely contributes to root hair curling and deformation in response to rhizobium inoculation.
CONCLUSION

This study represents one of the first large scale phosphoproteomic analyses that managed the complexity of sample size (nine time points, two conditions, and two tissues/cell types) and limited protein material (the root hair single cell) employing eight-plex iTRAQ and the LC platform specifically developed for sensitive and comprehensive phosphopeptide detection. Upon rhizobial inoculation, polar growth and deformation of the root hair cell, followed by infection thread development, initiates and then develops over a roughly 24 h period (69). However, signaling events within the root hair cell, in response to Nod factor addition, occur very rapidly, evidenced by changes in ion abundances in minutes, followed by apparent changes in the cellular cytoskeleton and gene expression within 3 h (69). The complexity of changes in the phosphoproteome identified in our study suggests that protein kinase cascades are likely an integral and important part of these cellular changes (Fig. 8). This study adds needed detail to our understanding of the means by which rhizobia infect legume roots to establish a nitrogen-fixing symbiosis and provides a wealth of additional gene and protein targets for future investigation.

Acknowledgments—We thank Samuel Purvine (Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory), Dr. Paul Grimsrud (Department of Biochemistry, University of Wisconsin-Madison) for technical support on phosphoproteomic analysis, Dr. Jianjiong Gao (Digital Biology Laboratory, University of Wisconsin-Madison) for technical support on phosphorylation motif discovery and the PDB database, Jun Wang and Sooyoung Franck for root hair isolation, Dr. Thang B. Dinh for graphic assistance, and Dr. Seth D. Findley and Dr. Katalin Toth for critical comments on the manuscript.

* This work was supported by the Plant Genome Program (DBI-0421620) from the National Science Foundation (to GS, LP-T, DX) and by funding from the United Soybean Board (to HN, GS, DX).

[1] S] This article contains supplemental Document, Figs. S1 to S5, and Tables S1 to S5.

To whom correspondence should be addressed: Dr. Gary Stacey, Divisions of Plant Sciences and Biochemistry, 2716 Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211. Tel.: 573-884-4752. E-mail: staceyg@missouri.edu.

The authors acknowledge that portions of this work were supported by the US Department of Energy (DOE) Office of Biological and Environmental Research and by the NIH National Center for Research Resources (RR018522). Work was performed in the Environmental Molecular Science Laboratory (EMSL), a DOE national scientific user facilities located on the campus of Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for DOE under contract DE-AC05-76RLO 1830.

REFERENCES

1. Gilroy, S., and Jones, D. L. (2000) Through form to function: root hair development and nutrient uptake. Trends Plant Sci. 5, 56–60
2. Bibikova, T., and Gilroy, S. (2003) Root hair development. J. Plant Growth Regul. 21, 383–415
3. Cárdenas, L., Holdaway-Clarke, T. L., Sánchez, F., Quinto, C., Feijó, J. A., Kunkei, J. G., and Hepler, P. K. (2000) Ion changes in legume root hairs responding to Nod factors. Plant Physiol. 123, 443–452
4. Hadri, A. E., and Bisseling, T. (1998). Responses of the plant to Nod factors. In: Späth, H. P., Dondorsh, A., and Hootykaus, P. J. J., eds. The Rhizobiales: Molecular Biology of Model Plant Associated Bacteria, pp. 403–416. Kluwer Academic Publishers, Dordrecht, The Netherlands
5. Oldroyd, G. E. D., and Downie, J. M. (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Ann. Rev. Plant Biol. 59, 519–546
6. Jones, M. A., Raymond, M. J., and Smimoff, N. (2006) Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in Arabidopsis. Plant J. 45, 83–100
7. Won, S. K., Lee, Y. J., Lee, H. Y., Heo, Y. K., Cho, M., and Cho, H. T. (2009) cisElement- and transcriptome-based screening of root hair-specific genes and their functional characterization in Arabidopsis. Plant Physiol. 150, 1473
8. Libault, M., Brechenmacher, L., Cheng, J., Xu, D., and Stacey, G. (2010) Root hair systems biology. Trends Plant Sci. 15, 641–650
9. Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T. H., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D., Cooper, B., and Stacey, G. (2009) Establishment of a protein reference map for soybean root hair cells. Plant Physiol. 149, 670–682
10. Libault, M., Farmer, A., Brechenmacher, L., Dnevich, J., Langley, R. J., Bilgin, D. D., Radman, O., Neece, D. J., Clough, S. J., May, G. D., and Stacey, G. (2010) Complete transcriptome of soybean root hair cell, a single cell model, and its alteration in response to Bradyrhizobium japonicum infection. Plant Physiol. 152, 541–552
11. Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., He, J., Xu, D., May, G., and Stacey, G. (2010) An integrated transcriptome atlas of the crop model Glycine max and its use in comparative analyses in plants. Plant J. 63, 86–99
12. Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczygowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J. (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425, 637–640
13. Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Gronlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., and Stougaard, J. (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425, 585–592
14. Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J., Szczygowski, K., and Parniske, M. (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417, 959–962
15. Endre, G., Kereszt, A., Kevei, Z., Mihácsa, S., Kaló, P., and Kiss, G. B. (2000) A receptor kinase gene regulating symbiotic nodule development. Nature 417, 962–966
16. Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulková, O., Duc, G., Journet, E. P., Ané, J. M., Lauber, E., Bisseling, T., Dénarié, J., Rosenberg, C., and Debélèf, F. (2004) A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303, 1361–1364
17. Mitrà, R. M., Gleason, C. A., Edwards, A., Hadfield, J., Downie, J. A., Oldroyd, G. E., and Long, S. R. (2004) A CA2+ and calmodulin-dependent kinase required for symbiotic nodule development: gene identification by transcript-based cloning. Proc. Natl. Acad. Sci. U.S.A. 101, 4701–4705
18. Yemets, A., Sheremet, Y., Vissenberg, K., Van Orden, J., Verbeelen, J. P., and Blume, Y. B. (2008) Effects of tyrosine kinase and phosphatase inhibitors on microtubules in Arabidopsis root cells. Cell Biol. Int. 32, 630–637
19. Lee, S. H., and Cho, H. T. (2006) PINOID positively regulates auxin efflux in Arabidopsis root hair cells and tobacco cells. Plant Cell 18, 1604–1616
20. Samaj, J., Ovečka, M., Hlavacka, A., Lecourieux, F., Mesiček, I., Lichtscheidl, I., Lenart, P., Salaj, J., Vollmann, D., Bögre, L., Baluska, F., and Hirt, H. (2002) Involvement of the mitogen-activated protein kinase SIMK in regulation of root hair tip growth. EMBO J. 21, 3296–3306
21. Yokota, K., Fukai, E., Madsen, L. H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Heid, M., Hossain, M. S., Szczygowski, K., Morieri, G., Oldroyd, G. E., Downie, J. A., Nielsen, M. W., Rusek, A. M., Sato, S., Tabata, S., James, E. K., Oyaizu, H., Sandal, N., and Stougaard, J. (2009) Rearrangement of actin cytoskeleton mediates invasion of Lotus japonicus roots by Mesorhizobium loti. Cell Plant 21, 267–284
22. Zhao, R., Ding, S., Shen, Y., Camp, II D. G., Liveyay, E. A., Udseth, H., and...
Quantitative Phosphoproteomics of Soybean Root Hair Cells

Smith, R. D. (2009) Automated metal-free multiple-column nanoLC for improved phosphoprotein analysis sensitivity and throughput. J. Chromatogr. B. 877, 663–670

23. Wang, J., Torres, M., Ganapathy, A., Thelen, J., DaQue, B. B., Mooney, B., Xu, D., and Stacey, G. (2005) Proteomic analysis of soybean root hair after infection by Bradyrhizobium japonicum. Mol. Plant-Microbe Interact. 18, 45–467

24. Ficarro, S. B., Adelmant, G., Tomar, M. N., Zhang, Y., Cheng, V. J., and Gygi, S. P. (2007) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. 25, 1285–1292

25. Prutt, K. D., Tatusova, T., Klimek, W., and Maglott, D. R. (2009) NCBI Reference Sequences: current status, policy, and new initiatives. Nucleic Acids Res. 37, D32–D36

26. de la Fuente, van Bentem, S., Anrather, D., Roitinger, E., Djamei, A., Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions Nat. Methods 8, 785–786

27. Krog, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Pattern recognition tools for transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580

28. Nihei, H., Brunak, S., Engbertsch, J., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6

29. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions Nat. Methods 8, 785–786

30. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580

31. Schlueter, J. A., Lin, Y. J., Schlueter, S. D., Vasylenko-Sanders, I. F., Deshpande, S., Yli, J., O'Brien, M., Roe, B. A., Nelson, R. T., Scheffler, B. E., Jackson, S. A., and Shoemaker, R. C. (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC Genomics 8, 330

32. Schmutz, J., Cannon, S. B., Schlueter, J. M., Metz, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellisten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharya, M. K., Sandhu, D., Villyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Sewell, J., Gill, N., Joshi, T., Labeau, M., Sethuraman, A., Zhang, X.-C., Xingzhi, K., Nguyen, H. T., Wing, A. R., Cregan, P., Schep, J., Grimwood, J., Roberts, D., Stacey, G., Shoemaker, R. C., and Jackson, S. A. (2010) Genome sequence of the paleopolyploid soybean. Nature 463, 178–183

33. Smith, J. C., and Figeys, D. (2008) Recent developments in mass spectrometry-based quantitative phosphoproteomics. Biochem. Cell Biol. 86, 137–148

34. Smith, J. C., Lambert, J. P., Elisma, F., and Figeys, D. (2007) Proteomics in 2005/2006: developments, applications and challenges. Anal. Chem. 79, 12534–12543

35. Thimm, O., Blasing, O., Giron, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L. A., Rhee, S. Y., and Stitt, M. (2004) MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939

36. Zhang, X. C., Wu, X., Findley, S., Wang, J., Libault, M., Nguyen, T. H., Cannon, B. S., and Stacey, G. (2007) Molecular evolution of Lysin motif-type receptor-like kinases in plants. Plant Physiol. 144, 623–636

37. Pauwels, L., Barbero, G. F., Geerinck, J., Tilleman, S., Grunewald, W., Perez, A. C., Chio, M., Bossche, R. V., Sewell, J., Gil, E., Garcia-Casado, G., Witters, E., Inzé, D., Long, J. A., Jaeger, G., Solano, R., and Goodwin, T. (2011)宁IAA connects the co-repressor TOPLESS to jasmonate signalling. Nature 464, 788–791

38. Wasternack, C. (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Annu. Bot. 100, 681–697

39. Delbarre, A., Muller, P., and Guern, J. (1998) Short-lived and phosphorylated proteins contribute to carrier-mediated efflux, but not to influx, of auxin in suspension-cultured tobacco cells. Plant Physiol. 116, 833–844

40. Wasson, A. P., Pellerone, F. I., and Mathesius, U. (2006) Silencing the flavonoid pathway in Medicago truncatula inhibit root nodule formation and prevents auxin transport regulation by rhizobia. Plant Cell 18, 1617–1629

41. Miller, D. D., de Ruijter, N. C. A., and Emons, A. M. C. (1997) From signal to form: aspects of the cytoskeleton-plasma membrane-cell wall continuum in root hair tips. J. Exp. Bot. 48, 1881–1896

42. Hepler, P. K., Vidali, L., and Cheng, A. (2001) Polarization of cell growth in higher plants. Annu. Rev. Cell Dev. Biol. 17, 159–187

43. de Ruijter, N. C. A., Rook, M. B., Bisseling, T., and Emons, A. M. C. (1998) Lipochito-oligosaccharides re-initiate root hair tip growth in Vicia sativa with high calcium and spectrin-like antigen at the tip. Plant J. 13, 341–350

44. Esseling, J. J., Lhuissier, F. G., and Emons, A. M. C. (2003) Nod factor-induced root hair curling: continuous polar growth towards the point of Nod factor application. Plant Physiol. 132, 1982–1988

45. Morre, D. J., and Van der Woude, W. J. (1974) Origin and growth of cell surface components. In: Hay, E., King, T. J., and Papakonstantinou, J., eds., Macromolecules Regulating Growth and Development, pp. 81–111, Academic Press, New York

46. Campanoni, P., and Blatt, M. R. (2007) Membrane trafficking and polar growth in root hairs and pollen tubes. J. Exp. Bot. 58, 65–74

47. Fu, Y., Li, H., and Yang, Z. (2002) The ROP2 GTPase controls the formation of cortical fine F-actin and the early phase of directional cell expansion during Arabidopsis organogenesis. Plant Cell 14, 777–794

48. Yang, Z. (2002) Small GTPases: Versatile signaling switches in plants. Plant Cell 14, S375–S388

49. Miller, D. D., de Ruijter, N. C. A., Emons, A. M. C. (1999) The role of actin in root hair morphogenesis: studies with lipochito-oligosaccharides as a growth stimulator and cytochalasin as an actin perturbing drug. Plant J. 17, 141–154

50. Hile, P., Vartainen, M. K., and Lappalainen, P. (2004) Regulation of the actin cytoskeleton by PI(4,5)P-2 and PI(3,4,5)P-3. Curr. Top. Microbiol. 282, 117–163

51. Saarkangas, J., Zhao, H., and Lappalainen, P. (2010) Regulation of the actin cytoskeleton-plasma membrane interplay by phosphoinositides. Plant Physiol. 150, 259–289

52. D’Souza-Schorey C., and Chavrier P. (2004) ARF proteins: roles in membrane traffic and beyond. Nat. Rev. Mol. Cell Biol. 7, 347–358

53. Gillingham A. K., and Munro S. (2007) The small G proteins of the Arf family and their regulators. Annu. Rev. Cell Dev. Biol. 23, 579–611

54. Curwin, A. J., Fairn, G. D., and McMasters, C. R. (2009) Phospholipid
transfer protein Sec14 is required for trafficking from endosomes and regulates distinct trans-Golgi export pathways. J. Biol. Chem. 284, 7364–7375
65. Stolarczyk, E. I., Reiling, C. J., and Paumi, C. M. (2011) Regulation of ABC Transporter Function via Phosphorylation by Protein Kinases. Curr. Pharm. Biotechnol. 12, 621–635
66. Paumi, C. M., Chuk, M., Chevelev, I., Stagljar, I., and Michaelis, S. (2008) Negative regulation of the yeast ABC transporter Ycf1p by phosphorylation within its N-terminal extension. J. Biol. Chem. 283, 27079–27088
67. Galway, M. E. (2006) Root hair cell walls: filling in the framework. Can. J. Bot. 84, 613–621
68. Cheung, A.Y., and Wu, H. M. (2008) Structural and signaling networks for the polar cell growth machinery in pollen tubes. Annu. Rev. Plant Biol. 59, 547–572
69. Day, R. B., Loh, J. T., Cohn, J., and Stacey, G. (1999) Chapter 22. Signal exchange involved in the establishment of the Bradyrhizobium-legume symbiosis. In: Triplett, E. (ed.), Prokaryotic Nitrogen Fixation: A Model System for the Analysis of a Biological Process, Horizon Scientific Press, Norfolk, England