Into the unknown: Using systems biology approaches in exploring new functions of plant receptor kinases

Waltraud Schulze

Department of Plant Systems Biology

systembiologie.uni-hohenheim.de
Plants are important

- Plants can adapt to almost all climates (latitude / altitude)
- Economic value: **Food, Feed, Aesthetic and Recreational**
- Influence on word climate

We would like to understand more about how plants function and how they adapt to a changing environment on the molecular level.

Image source: www.wikipedia.com
Image source: www.kawakarpo.de
Plant model: Challenging anatomy and biochemistry

- **Cell wall**: harsh extraction methods necessary with mechanical grinding.
- **Large central vacuole**: low protein yield, acidic, secondary metabolites
- **Secondary metabolites**: hydrophobic and interfere with protein purification
- **In green tissue**: Rubisco as highest abundant protein can mask detection of low-abundant proteins

Image source: www.wikipedia.com
Signaling pathways and mutants

- “One gene” resulting in “one phenotype” relationship was used to study signaling pathways.
- “Forward genetics”: Mutant phenotype is used to find underlying gene function.

**Plant hormone signaling pathways**

- **Brassinosteroid pathway**
  - WT
  - bri1-5
- **Ethylene pathway**
  - WT (etr1-1)

Gene functions and phenotypes in Arabidopsis

- About 27416 gene models in Arabidopsis.
- Up to now, 2400 genes in Arabidopsis are associated with a loss-of-function phenotype.
- Today, mainly “reverse-genetic” approaches are used to study gene functions.

Towards a „network biology“

Loss of protein function can be buffered by cellular network, leading to subtle phenotypes

Understanding protein function within the network requires:

- Parts list (measure proteins in the cell)
- Protein-Protein interactions, Protein signaling interactions

Network robustness and dynamics through „Proteoforms“

- Functional complexity is not achieved by many genes, but rather by a huge number of protein variations (Proteoforms).
- Proteoforms can result from combinations of:
  - SNPs
  - Splice variants
  - Posttranslational modifications


Image source: www.thermo.com
Importance of posttranslational modifications

- Posttranslational modifications serve multiple functions in protein-protein interaction (PIN) networks and protein signaling networks (PSN).
- Modifications and specific protein domains often form interaction sites of two proteins.

Kinases and phosphatases

- Phosphorylation is one of the most widespread posttranslational modifications (PTM).
- Almost all cellular processes are regulated by one or more phosphorylation or dephosphorylation events.
- Phosphorylation can be activating or inactivating.
- We still do not fully understand the functional context for most of the plant kinases.

Image source: www.thermo.com
Arabidopsis kinome

Cytosolic kinases

Receptor kinases

Monika Zulawski, PhD-student; Gunnar Schulze MSc-student
Membrane proteins and nutrient signaling

Carbon (sucrose)

Nitrogen (NO3, NH4)

Dynamics of Phosphorylation

Protein Complexes and Membrane Microdomains

Modification-Dependent Protein-Protein Interactions

Recognition, Uptake across membrane

Recognition, Signal transduction

Regulation of transcription, Regulation of protein activity
Quantitative proteomics strategies

**15N-Metabolic Labeling: Data Extraction Workflow**

- Labeled and unlabeled peptide form co-elute in reversed-phase liquid chromatography.
- Mass shift is dependent on the amino acid sequence.
- Identification of the peptide is necessary before the peptide pairs of labeled and unlabeled form can be quantified.

---


---

**Miancimu, 6054m, unclimbed**
Peptide Correlation Profiling: Data extraction workflow

Base peak chromatogram

Sample 1

Sample 2

Correlation of retention times

Retention time sample 1 [minutes]

Retention time sample 2 [minutes]

Retention time peak alignment increases quantitative coverage in label-free approaches.


Kawagebo, 6740m, unclimbed
Experimental setup: Nutrient starvation and resupply

Full medium → starvation → Re-supply of nutrient

NO$_3$/NH$_4$ → sucrose

Biological Experiment

starved → 3 minutes → 5 minutes → 10 minutes → 30 minutes

Sample Preparation

Plasma membranes right side out → Flip 0.01% Brij58 → Brij 58 treated vesicles inside-out → Digest Trypsin → LC-MS/MS

Identification

Quantitation

Nutrient-dependent phosphorylation response classes

- K-means clustering of phosphorylation time profiles.
Protein functions represented in phosphorylation responses

- Early responses (up to 5 min) involves receptor kinases, cell organization, transcription factors, GPI-anchored membrane proteins.
- Late responses (after 5 min) involve protein degradation and synthesis, second messenger signaling and central metabolism, such as enzymes.

Legend:
- transport
- cell organization
- protein degradation/synthesis
- GPI-anchored plasma membrane
- signaling
- transcription factors
- glycolysis & metabolism
- hormone metabolism

Membrane proteins and nutrient signaling

Carbon (sucrose)
Nitrogen (NO3, NH4)

Phosphorylation-Dependent Protein-Protein Interactions
Dynamics of Phosphorylation
Protein Complexes and Membrane Microdomains
Recognition, Uptake across membrane
Recognition, Signal transduction
Regulation of transcription, Regulation of protein activity

Experimental: Functional analysis of candidate proteins
Computational: Analysis of phosphorylation networks
PhosPhAt – Arabidopsis protein phosphorylation site database and phosphorylation site predictor

http://phosphat.mpimp-golm.mpg.de

Kinase target search in PhosPhAt

<table>
<thead>
<tr>
<th>Kinase: AT1G68830.1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G68830.1</td>
<td>kinase</td>
<td>other kinase</td>
<td>phosphorylation</td>
<td>PS</td>
<td>AT3G47470.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kinase: AT2G43790.1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>phosphorylation</td>
<td>stress.biotic,respiratory b...</td>
<td>AT3G25250.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>phosphorylation</td>
<td>stress.biotic,respiratory b...</td>
<td>AT2G47060.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>pathway</td>
<td>hormone,jasmonate</td>
<td>AT5G53210.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>phosphorylation</td>
<td>stress.biotic,plant defence</td>
<td>AT5G54430.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>phosphorylation</td>
<td>development, stomata</td>
<td>AT5G20770.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>activation</td>
<td>hormone,jasmonate</td>
<td>AT3G20770.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>activation</td>
<td>hormone,salicylic acid</td>
<td>AT3G20770.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>pathway</td>
<td>stress.abiotic</td>
<td>AT4G01250.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>activation</td>
<td>stress.biotic,plant defence</td>
<td>AT4G23550.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>activation</td>
<td>stress.biotic,plant defence</td>
<td>AT4G23550.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>pathway</td>
<td>stress.biotic,plant defence</td>
<td>AT4G23550.1</td>
</tr>
<tr>
<td>AT2G43790.1</td>
<td>kinase</td>
<td>MAP K</td>
<td>pathway</td>
<td>stress.biotic,plant defence</td>
<td>AT4G23550.1</td>
</tr>
</tbody>
</table>

Monika Zulawski, PhD-student

Integrated phosphorylation site predictor

*Integrated phosphorylation site predictor*

- Based on support-vector machine.
- Feature Vector evaluates information on local sequence of amino acids, amino acid chemical and physical properties as well as structural disorder indices.

Different natural variants of Arabidopsis have been characterized with different phenotypes.

Sequencing efforts of these accessions provides a basis for new system-wide analyses to address evolutionary context of adaptation to environments.
Gain and loss of phosphorylation sites by SNPs

Loss of a psite by a non-synonymous Single Nucleotide Polymorphism (SNP)

A – Adenin
cC - Cytosin
tT – Thymin
gG – Guanin

Genotype 1 of Arabidopsis: T C T → Ser
Genotype 2 of Arabidopsis: G C T → Ala → loss of psite

➤ Adaptation to different environments?

➤ Which proteins are affected by gain/loss of psites?

➤ Are particular protein functions over-represented in the proteins affected by gain/loss of psites?
Mapping of nsSNPs to phosphorylation sites

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of non-redundant SNPs in this study</th>
<th>Number of non-redundant SNPs mapping onto cDNAs</th>
<th>Number of non-redundant SNPs mapping onto CDS</th>
<th>Number of non-redundant SNPs causing at least one non-synonymous substitution</th>
<th>Number of non-redundant SNPs always causing synonymous substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nordborg</td>
<td>20667</td>
<td>9251</td>
<td>8023</td>
<td>4047</td>
<td>3975</td>
</tr>
<tr>
<td>Clark</td>
<td>637522</td>
<td>263718</td>
<td>227497</td>
<td>109709</td>
<td>117788</td>
</tr>
<tr>
<td>Ossowskii</td>
<td>860154</td>
<td>220984</td>
<td>174559</td>
<td>84400</td>
<td>90159</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1247284</td>
<td>382770</td>
<td>315039</td>
<td>156034</td>
<td>159004</td>
</tr>
</tbody>
</table>

156034 nsSNPs

- 7178 experimental psites
- 86 loss sites (86 proteins)
- 1114 loss sites
- 1148 gain sites
- 75,396 hc predicted psites

### Table 1: Arabidopsis Proteins with a Loss of Psite by nsSNP

<table>
<thead>
<tr>
<th>AGI</th>
<th>Protein position</th>
<th>Status</th>
<th>Accession/substitution</th>
<th>TAIR7 function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G01550.1</td>
<td>337 (pS)</td>
<td>C24(A); Got-7(A); Bur-6(G)</td>
<td>BPS1 (BYPASS 1)</td>
<td></td>
</tr>
<tr>
<td>AT1G02850.1</td>
<td>402 (pS)</td>
<td>Bay-0(A); Br-0(A); Est-0(A); Nfa-8(A); Tam-2(A); Tsu-1(A); Van-0(A)</td>
<td>Glycosyl hydrolase family 1 protein</td>
<td></td>
</tr>
<tr>
<td>AT1G13410.1</td>
<td>216 (pT)</td>
<td>Br-0(N); C24(N); Cvi-0(N); Est-1(N); Nfa-8(N); Rrs-10(N); Tsu-1(N); Shal(N)</td>
<td>Catalytic</td>
<td></td>
</tr>
<tr>
<td>AT1G19410.1</td>
<td>6 (pT)</td>
<td>Bur-0(S)</td>
<td>Similar to F-box family protein</td>
<td></td>
</tr>
<tr>
<td>AT1G22060.1</td>
<td>261 (pS)</td>
<td>Shal(N)</td>
<td>Similar to F-box family protein</td>
<td></td>
</tr>
<tr>
<td>AT1G22170.1</td>
<td>56 (pY)</td>
<td>Bur-0(F)</td>
<td>Phosphoglycerate/bisphosphoglycerate mutase family protein</td>
<td></td>
</tr>
<tr>
<td>AT1G24250.1</td>
<td>45 (pT)</td>
<td>Bur-0(A); Shal(A)</td>
<td>Paired amphiphatic helix repeat-containing protein</td>
<td></td>
</tr>
<tr>
<td>AT1G33530.1</td>
<td>2 (pT)</td>
<td>Tam-2(A); Tsu-1(A)</td>
<td>F-box family protein</td>
<td></td>
</tr>
<tr>
<td>AT1G44800.1</td>
<td>341 (pS)</td>
<td>Bar-0(N); Est-1(Y); Fes-0(Y); Got-7(Y); Let-1(Y); Nfa-8(Y); Rrs-10(Y); Rss-7(Y); Tam-2(Y); Tsu-1(Y); Van-0(Y)</td>
<td>Nodulin MtN21 family protein</td>
<td></td>
</tr>
<tr>
<td>AT1G52920.1</td>
<td>37 (pS)</td>
<td>Bur(P); C24(P); Est-1(P); Lov-5(P); Nfa-8(P); Rrs-7(P); Tam-2(P); Tsu-1(P); Van-0(P);</td>
<td>Catalytic</td>
<td></td>
</tr>
<tr>
<td>AT1G62330.1</td>
<td>500 (pS)</td>
<td>Est-1(G); Rss-10(G); Tam-1(N)</td>
<td>Similar to unknown protein</td>
<td></td>
</tr>
<tr>
<td>AT1G63870.1</td>
<td>856 (pT)</td>
<td>Shal(A)</td>
<td>Disease resistance protein (TIR-NBS-LRR class); patatin</td>
<td></td>
</tr>
<tr>
<td>AT1G63900.1</td>
<td>278 (pT)</td>
<td>Bur-0(A); Tsu-1(A)</td>
<td>Zinc-finger (C3HC4-type RING finger family) protein</td>
<td></td>
</tr>
<tr>
<td>AT1G64580.1</td>
<td>539 (pS)</td>
<td>Ler-1(P); Tsu-1(P)</td>
<td>Pentatricopeptide (PPR) repeat-containing protein</td>
<td></td>
</tr>
<tr>
<td>AT1G65060.1</td>
<td>155 (pS)</td>
<td>C24(T)</td>
<td>4CL3 (4-coumarate/CoA ligase 3); 4-coumarate-CoA ligase</td>
<td></td>
</tr>
<tr>
<td>AT1G65260.1</td>
<td>173 (pT)</td>
<td>Bay-0(S); Tsu-1(S)</td>
<td>ATTC4 (PLASTID TRANSCRIPTIONALLY ACTIVE)</td>
<td></td>
</tr>
<tr>
<td>AT1G68760.1</td>
<td>100 (pS)</td>
<td>Bur-0(T); Cvi-0(T); Est-0(T); Fes-0(T); Got-7(T); Tsu-1(T)</td>
<td>ATNUD1 (Arabidopsis thaliana Nudix hydrolase homolog 1); dihydroxyacetone pyrophosphatase pyrophosphohydrolase</td>
<td></td>
</tr>
<tr>
<td>AT1G71710.1</td>
<td>306 (pS)</td>
<td>C24(C)</td>
<td>Inositol polyphosphate 5-phosphatase putative</td>
<td></td>
</tr>
<tr>
<td>AT1G78360.1</td>
<td>197 (pS)</td>
<td>Bay-0(C)</td>
<td>ATGSTU21 (Arabidopsis thaliana Glutathione S-transferase (class tau) 21-glutathione transferase)</td>
<td></td>
</tr>
<tr>
<td>AT1G79850.1</td>
<td>4 (pS)</td>
<td>Bay-0(T); Bot-4(T); Est-1(T); Rss-7(T)</td>
<td>RPS17 (ribosomal protein S17); structural constituent of ribosome</td>
<td></td>
</tr>
<tr>
<td>AT2G02160.1</td>
<td>552 (pT)</td>
<td>C24(M); Tam-2(M)</td>
<td>Zinc finger (CUCCH-type) family protein</td>
<td></td>
</tr>
<tr>
<td>AT2G03110.1</td>
<td>97 (pT)</td>
<td>Bur-0(K); Tsu-1(K)</td>
<td>Nucleic acid binding</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Additional Receptor Proteins

<table>
<thead>
<tr>
<th>AGI</th>
<th>Protein position</th>
<th>Status</th>
<th>Accession/substitution</th>
<th>TAIR7 function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G20710.1</td>
<td>301 (pY)</td>
<td>Bay-0(C); Bot-4(C); Br-0(C); Bur-0(C); C24(C); Fes-0(C); Let-1(C); Lov-5(C); Nfa-8(C); Rrs-7(C); Shal(N)</td>
<td>Pentatricopeptide (PPR) repeat-containing protein</td>
<td></td>
</tr>
<tr>
<td>AT2G23360.1</td>
<td>600 (pT)</td>
<td>Est-0(M)</td>
<td>Transport protein-related</td>
<td></td>
</tr>
<tr>
<td>AT2G27120.1</td>
<td>1000 (pS)</td>
<td>Ag-0(R); CII-5(R); Cvi-0(R); HR-10(R); NFA-10(R); NFA-4(R); Pus-0(R); Sqa-1(R); Tam-2(R); Tam-27(R); ULE-5(R); Var-2-1(R); Var-2-6(R)</td>
<td>POL2B/TIL2 (TILTED2); DNA-directed DNA polymerase</td>
<td></td>
</tr>
</tbody>
</table>

**Receptor proteins are overrepresented**

Functional over- and underrepresentation in proteins with phosphorylation sites affected by nsSNPs

- Receptor proteins and other proteins involved in signaling and stress response are over-represented in proteins affected by gain or loss of phosphorylation sites.
- Regulation in metabolism is more conserved.

Goal: Time-resolved phosphorylation network

Aim: Reconstruct phosphorylation/signaling network from time course data.
- Evaluation of different methods for network reconstruction.
- Study network properties (degree, path lengths etc.).

Validation of results
- Analysis of network properties in selected k.o. mutants / site-directed mutants.
- Mapping to protein-protein interactions to network.
Mapping phosphoproteins to protein-protein interaction network

http://bioinfo.esalq.usp.br/atpin/atpin.pl

- Phosphoproteins particularly map to hub-proteins with high degree.
- Average degree of phosphoproteins is higher than that of non-phosphoproteins.

Size represents the degree of each node

Cyan nodes represent phosphoproteins
Degree distribution of modified proteins

- Proteins with phosphorylation and acetylation sites have highest degree in protein-protein interaction networks.
- Proteins with N-glycosylation are less central in PINs.

GuangYou Duan; PhD-student
Protein-protein interactions and subcellular location

<table>
<thead>
<tr>
<th>Protein</th>
<th>nucleus</th>
<th>golgi</th>
<th>mitochondrion</th>
<th>endoplasmic reticulum</th>
<th>cytosol</th>
<th>unclear</th>
<th>extracellular</th>
<th>vacuole</th>
<th>plasma membrane</th>
<th>plastid</th>
<th>peroxisome</th>
<th>cytoskeleton</th>
<th>cell plate</th>
<th>endosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>nucleus</td>
<td>0.349309</td>
<td>0.342722</td>
<td>0.26194514</td>
<td>0.122290362</td>
<td>0.358075</td>
<td>0.215426</td>
<td>0.16257</td>
<td>0.336244</td>
<td>0.201606</td>
<td>0.395944</td>
<td>0.260160</td>
<td>0.336504</td>
<td>0.6549022</td>
<td>1.203782163</td>
</tr>
<tr>
<td>golgi</td>
<td>0.342722</td>
<td>0.446403</td>
<td>0.40355089</td>
<td>0.438802186</td>
<td>0.485096</td>
<td>0.541328</td>
<td>0.355661</td>
<td>1.079857</td>
<td>0.851003</td>
<td>0.368099</td>
<td>0.365264</td>
<td>1.178952594</td>
<td>2.69918094</td>
<td>8.918295893</td>
</tr>
<tr>
<td>mitochondrion</td>
<td>0.261945</td>
<td>0.403565</td>
<td>0.202292536</td>
<td>0.165996525</td>
<td>0.340957</td>
<td>0.253059</td>
<td>0.19602</td>
<td>0.406483</td>
<td>0.406264</td>
<td>0.267595</td>
<td>0.328373</td>
<td>0.752471592</td>
<td>0.45055459</td>
<td>1.616490307</td>
</tr>
<tr>
<td>endoplasmic reticulum</td>
<td>0.13228</td>
<td>0.433902</td>
<td>0.185998952</td>
<td>0.452001381</td>
<td>0.232256</td>
<td>0.42089</td>
<td>0.308875</td>
<td>0.475452</td>
<td>0.333414</td>
<td>0.183593</td>
<td>0.201154</td>
<td>0.250951077</td>
<td>0.35920875</td>
<td>1.722850744</td>
</tr>
<tr>
<td>cytosol</td>
<td>0.350075</td>
<td>0.48506</td>
<td>0.340956664</td>
<td>0.202253685</td>
<td>0.408513</td>
<td>0.342719</td>
<td>0.249306</td>
<td>0.554835</td>
<td>0.56650</td>
<td>0.339976</td>
<td>0.472931</td>
<td>1.212644727</td>
<td>0.73933001</td>
<td>1.966019300</td>
</tr>
<tr>
<td>unclear</td>
<td>0.215426</td>
<td>0.541328</td>
<td>0.253552555</td>
<td>0.420885565</td>
<td>0.342715</td>
<td>0.622917</td>
<td>0.33204</td>
<td>0.577629</td>
<td>0.463857</td>
<td>0.262902</td>
<td>0.373217</td>
<td>0.644650308</td>
<td>0.61530889</td>
<td>2.227897625</td>
</tr>
<tr>
<td>extracellular</td>
<td>0.16257</td>
<td>0.35541</td>
<td>0.19501973</td>
<td>0.308926444</td>
<td>0.249396</td>
<td>0.3204</td>
<td>0.292965</td>
<td>0.437020</td>
<td>0.375287</td>
<td>0.199793</td>
<td>0.254712</td>
<td>0.455004619</td>
<td>0.22613562</td>
<td>1.342322703</td>
</tr>
<tr>
<td>vacuole</td>
<td>0.333624</td>
<td>1.079857</td>
<td>0.408482748</td>
<td>0.475467127</td>
<td>0.554835</td>
<td>0.577629</td>
<td>0.437020</td>
<td>1.146297</td>
<td>0.821964</td>
<td>0.399219</td>
<td>0.502615</td>
<td>1.518688063</td>
<td>0.54098821</td>
<td>4.510375860</td>
</tr>
<tr>
<td>plasma membrane</td>
<td>0.39544</td>
<td>0.051003</td>
<td>0.436260497</td>
<td>0.302414196</td>
<td>0.56636</td>
<td>0.460527</td>
<td>0.375207</td>
<td>0.921994</td>
<td>0.501427</td>
<td>0.427192</td>
<td>0.503925</td>
<td>1.203672707</td>
<td>0.56002793</td>
<td>3.392709932</td>
</tr>
<tr>
<td>plastid</td>
<td>0.260158</td>
<td>0.368059</td>
<td>0.267958001</td>
<td>0.163590093</td>
<td>0.359976</td>
<td>0.28902</td>
<td>0.193793</td>
<td>0.395219</td>
<td>0.427153</td>
<td>0.280379</td>
<td>0.364594</td>
<td>0.749454338</td>
<td>0.48344938</td>
<td>1.651592607</td>
</tr>
<tr>
<td>peroxisome</td>
<td>0.336</td>
<td>0.345264</td>
<td>0.320372175</td>
<td>0.241153074</td>
<td>0.471931</td>
<td>0.373217</td>
<td>0.254712</td>
<td>0.502615</td>
<td>0.503925</td>
<td>0.349694</td>
<td>0.865399</td>
<td>1.24673065</td>
<td>0.38934150</td>
<td>1.45302925</td>
</tr>
<tr>
<td>cytoskeleton</td>
<td>1.074094</td>
<td>1.178953</td>
<td>0.752471592</td>
<td>0.350951077</td>
<td>1.221646</td>
<td>0.644609</td>
<td>0.455805</td>
<td>1.518689</td>
<td>1.283674</td>
<td>0.794554</td>
<td>1.266781</td>
<td>28.45095692</td>
<td>6.92041522</td>
<td>8.921390574</td>
</tr>
<tr>
<td>cell plate</td>
<td>0.6549</td>
<td>2.699181</td>
<td>0.459654595</td>
<td>0.35920875</td>
<td>0.73033</td>
<td>0.615309</td>
<td>0.226136</td>
<td>0.946399</td>
<td>0.560026</td>
<td>0.403449</td>
<td>0.360342</td>
<td>6.250148522</td>
<td>55.36321209</td>
<td>10.055129717</td>
</tr>
<tr>
<td>endosome</td>
<td>1.203782</td>
<td>8.918257</td>
<td>1.616498387</td>
<td>1.722850744</td>
<td>1.966815</td>
<td>2.227988</td>
<td>1.343823</td>
<td>4.510374</td>
<td>3.338279</td>
<td>1.651591</td>
<td>1.453029</td>
<td>8.021390374</td>
<td>10.6951872</td>
<td>55.11570248</td>
</tr>
</tbody>
</table>

- Subcellular locations of proteins are used to assess best method for network reconstruction.
- Subcellular location can be used for weighting predicted interactions.
- Partial correlation method gave the best results in our dataset.
Reconstructed network topology

- **Initiation layer**: Plasma membrane proteins
- **Processing layer**: highest overlap among treatments
- **Effector layer**: cytosolic proteins, nuclear proteins. MAPK-motifs
Time-resolved phosphorylation network

- Reconstructed network using Graphical Gaussian Models (GGM).
- Network neighbors: of a receptor kinase: sucrose transporters; aquaporins; phosphatase 2C.
Membrane proteins and nutrient signaling

Carbon (sucrose)  Nitrogen (NO3, NH4)

Recognition, Uptake across membrane

Protein Complexes and Membrane Microdomains

Phosphorylation-Dependent Protein-Protein Interactions

Dynamics of Phosphorylation

Computational: Analysis of phosphorylation networks

Experimental: Functional analysis of candidate proteins

Recognition, Signal transduction

Regulation of transcription, Regulation of protein activity

Phosphorylation - Dependent Protein-Protein Interactions

Dynamics of Phosphorylation

Computational: Analysis of phosphorylation networks

Experimental: Functional analysis of candidate proteins
SIRK1: Sucrose-Induced Receptor Kinase

A

SIRK1 (3436bp)

5' ATG  

Intron

Exon

TGA 3'

F1 LP

R1

T-DNA

RP

LBb1.3

B

sirk1  Col-0

(bp)

800  

600

600  

400

LP/RP

LBB1.3/RP

C

Relative Expression (2^ΔCt)

Col-0  sirk1  35S::AtSIRK1-GFP  35S::AtSIRK1+44A-GFP  35S::AtSIRK1+140D-GFP

**  **  **  **
Functional characterization using phosphorylation site mutants

Transformation of sirk1 T-DNA insertional mutant with:
- SIRK cDNA
- SIRK1 with mutated Ser-744 to Asp-744 (D); phosphorylation mimic mutant
- SIRK1 with mutated Ser-744 to Ala-744 (A); phosphorylation null mutant
Luciferase-based kinase-activity assay shows sucrose-dependent quench of luminescence when expressing SIRK1 in *sirk1* background.

- SIRK1 is a protein kinase
- SIRK1 activity is increased by sucrose
SIRK1 interaction partners

SIRK1-GFP
SIRK1_{S744D}-GFP
SIRK1_{S744A}-GFP
GFP (control)

0 suc 3min suc
0 suc 3min suc
0 suc 3min suc
0 suc 3min suc

Microsome preparation

GFP-IP (Chromotek)

Data analysis:
MaxQuant
cRacker

LC-MS/MS

Xuna Wu, PhD-thesis
SIRK1 interaction partners

Over-representation of: Clathrins, coatomers

Over-representation of aquaporins: PIP2E, PIP2B, PIP2A, PIP1A, PIP1D, PIP3A

Common interaction partners with plasma membrane location include:
- Over-representation of aquaporins (6 isoforms)
- Stronger interactions with S744D:
  - Clathrins and coatomers involved in vesicle trafficking

Xuna Wu, PhD-thesis
Functional analysis using kinase mutants

Full nutrition → Sucrose Starvation → Sucrose Re-supply → Mass spectrometry

Col-0  sirk1  SIRK1  SIRK1 S744D  SIRK1 S744A

14 days → 2 days → 3 minutes

Protein extraction, Membrane preparation, TiO2
Altered phosphorylation patterns in kinase mutant *sirk1*

Several transport proteins show reduced sucrose-induced phosphorylation in *sirk1* mutant.
Altered phosphorylation particularly of transport proteins

<table>
<thead>
<tr>
<th>Function</th>
<th>Locus</th>
<th>Phosphopeptides</th>
<th>Bin</th>
<th>Log₂ fold (sp1 vs WT)</th>
<th>Description/Name</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport related</td>
<td>AT3G58730</td>
<td>GI(pS)INAAR</td>
<td>34.1.1.3</td>
<td>-1.11</td>
<td>VHA-D</td>
<td>Niittylä et al. 2007</td>
</tr>
<tr>
<td>Transport related</td>
<td>AT4G30190</td>
<td>GLDIETPSHYTV</td>
<td>34.1.2</td>
<td>-0.5</td>
<td>AHA2</td>
<td>ATPase</td>
</tr>
<tr>
<td>Transport related</td>
<td>AT3G01390</td>
<td>ME(pS)NRGQQGISIQQLAAEVEAQHIVNAAR</td>
<td>34.1</td>
<td>-0.65</td>
<td>VMA10</td>
<td></td>
</tr>
<tr>
<td>Transport related</td>
<td>AT5G57110</td>
<td>TSLLK(pS)SPGR</td>
<td>34.21</td>
<td>-0.69</td>
<td>ACA8</td>
<td></td>
</tr>
<tr>
<td>Transport related</td>
<td>AT1G57990</td>
<td>QTTAEGSANPEPDQIL (pS)PR</td>
<td>34.10</td>
<td>-1.08</td>
<td>ATPUP18</td>
<td></td>
</tr>
<tr>
<td>Transport related</td>
<td>AT5G27150</td>
<td>GFVPFVPG(pS)PTER</td>
<td>34.14</td>
<td>-0.72</td>
<td>NHX1</td>
<td>purine transporter</td>
</tr>
<tr>
<td>Transport related</td>
<td>AT2G37170</td>
<td>SLG(pS)FR(pS)AANV</td>
<td>34.19.1</td>
<td>-0.83</td>
<td>PIP2B</td>
<td>aquaporin</td>
</tr>
<tr>
<td>Transport related</td>
<td>AT4G35100</td>
<td>ALG(pS)FR(pS)NATN</td>
<td>34.19.1</td>
<td>-2.26</td>
<td>PIP3</td>
<td></td>
</tr>
<tr>
<td>Transport related</td>
<td>AT5G60660</td>
<td>ALGSFG(pS)FG(pS)FR</td>
<td>34.19.1</td>
<td>-1.33</td>
<td>PIP2;4/PIP2F</td>
<td></td>
</tr>
<tr>
<td>Transport related</td>
<td>AT1G75220</td>
<td>(pS)FRDDNNEEAR</td>
<td>34.2</td>
<td>-1.78</td>
<td>unknown</td>
<td>Sugar transporter</td>
</tr>
<tr>
<td>Transport related</td>
<td>AT3G48740</td>
<td>LGTVS(pS)PEPISVVR</td>
<td>33.99</td>
<td>-0.91</td>
<td>ATSWEET 11</td>
<td>Sucrose transporter</td>
</tr>
<tr>
<td>Signaling</td>
<td>AT3G02880</td>
<td>LIEEVSHSSG(pS)PNPVSND</td>
<td>30.2.3</td>
<td>-0.54</td>
<td>LRR-RLK</td>
<td>receptor kinase</td>
</tr>
<tr>
<td>Signaling</td>
<td>AT2G43130</td>
<td>(pS)DDDERGEEEYLFK</td>
<td>30.5</td>
<td>-0.5</td>
<td>ARA4</td>
<td>G-protein</td>
</tr>
</tbody>
</table>

Does SIRK1 directly modulate transport proteins?

Xuna Wu, PhD-thesis
Validation: Protoplast swelling assay

Sucrose-starved seedlings → Protoplast Isolation (500 mM mannitol) → Protoplast treatment with 30 mM sucrase → Protoplast swelling observation

Low osmolarity: 500 mM Mannitol (ES-M500)
High osmolarity: 350 mM Mannitol (ES-M350)

High osmolarity:
- 500 mM Mannitol (ES-M500)
- 470 mM Mannitol/30 mM Sucrose (ES470M_30S)

Low osmolarity:
- 350 mM Mannitol (ES-M350)
- 320 mM Mannitol/30 mM Sucrose (ES320M_30S)

Measurement of Water flux density based on the volume change of protoplast per unit surface area
Volume flow density $j_v = \Delta V / (\Delta t * S_0)$

Protoplast swelling assay with wild type, sirk1 mutant and transformants (SIRK1, SIRK1A, SIRK1D)
Sucrose or mannitol as solute

Xuna Wu, PhD-thesis
SIRK1 modulates water flux in sucrose-dependent manner

- Sucrose-induced swelling is significantly reduced in sirk1
- Sucrose-induced swelling is restored in SIRK1-GFP-line
- No difference between lines is observed for mannitol treatment
SIRK1 S744 is involved in receptor internalization

- Sucrose induces SIRK1 internalization within 15 minutes
- Mutation S744A displays reduced SIRK1 internalization
- Mutation S744D results in sucrose-independent SIRK1 internalization.
Functional model for SIRK1

S744A  |  SIRK1  |  S744D
---|---|---
clathrin  |  aquaporin  |  clathrin
memorin  |  SIRK1  |  SIRK1
AT3G02880  |  SWEET11  |  clathrin

sucrose starvation

sucrose resupply 3min

sucrose resupply 10min

SIRK1 phosphorylation declines
SIRK1 dimerization becomes less
Phosphorylation activity of targets goes down
Phosphorylated SIRK1 forms are internalized
Summary

- Dynamic analysis of protein modification allows definition of edges in signaling networks.
- Network model suggests relevant sub-networks relevant under particular biological context (cue-signal response compendiums).
- Kinase mutants and phosphorylation site mutants are important tools for characterization of regulatory interactions in protein signaling networks.
- For rapid adaptations to changing external conditions, direct regulations of membrane transporters by membrane bound kinases become evident.
Signaling Proteomics Group

Slava Braginets (Programmer)
Sylwia Kierszniowska (PhD, PostDoc)
Heidi Pertl-Obermeyer (PostDoc)
Witold Szymanski (PhD)
Karen Vester (Bsc)
XuNa Wu (PhD, PostDoc)
Kerstin Zander (TA)
Monika Zulawski (PhD)
Wolfgang Engelsberger (PhD)
John Rohloff (BSc)
Robert Schmidt (Programmer)
Margitta Schümann (TA)
Alois Schweighofer (PostDoc)
Dmitri Schmidt (Bsc, MSc)
Natalie Arndt (Dipl)
Borjana Arsova (PostDoc)
Vivian Schüler (Bsc)
Gunnar Schulze (Msc)
Henrik Zauber (Dipl, PhD)

Collaborators:
Dirk Walther, MPI-MPIMP
Staffan Persson, MPI-MPIMP
Stefanie Hartmann, UP

“All true wisdom is only to be found far from the dwellings of men, in the great solitudes; and it can only be attained through suffering.”
(Igjagarjuk, Greenland)
**15N-Metabolic labeling: Data extraction workflow**

- Labeled and unlabeled peptide form co-elute in reversed-phase liquid chromatography.
- Mass shift is dependent on the amino acid sequence.
- Identification of the peptide is necessary before the peptide pairs of labeled and unlabeled form can be quantified.

---

Peptide Correlation Profiling: Data extraction workflow

A. Base peak chromatogram
B. Correlation of retention times
C. Extracted ion chromatogram
D. MS2 fragment spectrum
E. MS full scan

- Retention time peak alignment increases quantitative coverage in label-free approaches.

Summary

- Dynamic analysis of protein modification allows definition of edges in signaling networks.
- Network model suggests relevant sub-networks relevant under particular biological context (cue-signal response compendium).
- Kinase mutants and phosphorylation site mutants are important tools for characterization of regulatory interactions in protein signaling networks.
- For rapid adaptations to changing external conditions, direct regulations of membrane transporters by membrane bound kinases become evident.

To be done:
- Integration of \textit{sirk1} data set into the dynamic network.
- Targeted analysis of phosphorylation stoichiometry in CSR compendiums
Plans and Ideas

Research:
- Protein-Protein Interaction Networks and Protein Signaling Networks: Use quantitative data for true modeling
- PINs and PSNs in natural variation (e.g. *Arabidopsis* accessions)
- Quantitation of phosphorylation stoichiometry
- Small molecule–Protein Interactions / protein-lipid interaction
- Dynamics of membrane microdomains in context of external perturbations

Teaching:
- Large-scale high-throughput methods
- Experimental design and data analysis in quantitative proteomics
- „Big data“ and networks (PINs, PSNs, CSRs)
- Primary metabolism and nutrient signaling in plants
- Plant signaling networks
Proteome space accessible for quantitation


- Instrumentation has and will be improved:
  - Scan speed.
  - Mass accuracy.
  - Sensitivity.
  - Selectivity.

- Analysis of accurate and full proteome coverage become feasible.
Towards a „network biology“

Networks consist of nodes and connecting edges.

- **Protein-protein interaction networks (PIN)**: undirected, direct physical interactions
- **Protein-signaling networks (PSN)**: directed, not equal to PINs.

Measured by: AP-MS Y2H cell biology

Measurement ?:
transient often, edges need to be inferred

- Protein interaction networks are undirected
- Protein-signaling networks are directed

Combined cellular network

Towards a „network biology“

Data sets for “network biology” need to fulfill these criteria:
- Complete datasets (all nodes and edges need to be measurable)
- Reproducible data sets
- Quantitative data
- Reasonable throughput

---

Towards a „network biology“


Diagram:
- Describing network wiring
  - Enzymatic assays
  - Enzyme-substrate prediction
  - Network reverse engineering
  - Y2H
  - Genetic interaction screen

- Dynamic network rewiring
  - AP MS
  - AP SRM
  - Perturbation analysis

- Correlating network wiring to phenotype
  - Kinase-selective enrichment
  - Profiling enzymatic activity
  - Projecting data on static networks

- Experimental
- Computational
Towards a „network biology“

- Essential genes are enriched for RNA and protein synthesis and protein modification, but deficient in transcription factors.
- Genes inducing morphological phenotypes are enriched in transcriptional regulation and signaling functions.
- Essential genes are more likely to be unique (no functional redundancy).

Quantitative proteomics workflow

- Raw data are being processed to allow spectra to be assigned to peptide sequences.
- Quantitative information is extracted based on MS or MS/MS spectra.
- Spectra with assigned peptide sequences are linked to quantitative information.
- Identification and quantitation usually is validated by a manual verification step.
- Averaged peptide and protein ratios are then subjected to statistic analysis and final evaluation.

Integration of different proteomic resources in a single portal:
- AtPeptide
- SUBA
- TAIR
- PhosPhAt
- PPDB
- ...
Reproducibility of experimental phosphorylation sites

- Most phosphorylation sites and most phosphoproteins have been identified only once (in one experiment).
- Only 2% of the phosphorylation sites have been identified by more than two different studies.

Arabidopsis kinome and gene duplications

Monika Zulawski, PhD-student
Gene duplications in different kinase families

Gene duplications in different kinase families

Monika Zulawski, PhD-student
Functional diversification upon gene duplication

+ phylogeny

- phylogeny

LRR_3

RLCK_9

LRR_8B

Monika Zulawski, PhD-student
Mass spectrometry based proteomics

Protein mixture

Separation by 2D-gel select and isolate spots

Identify proteins by peptide mass fingerprint

Complex peptide mixture separate by liquid chromatography

Identify proteins by tandem mass spectrometry
cRacker as automated and standardized data analysis tool

Henrik Zauber; PhD-thesis


- **cRacker**
- **automated**
- **standardized data analysis tool**

**cRacker produces graphic output and statistics:**
- **ANOVA, t-test:** presentation as volcano plots
- **K-means or hierarchical clustering**
- **Over-representation test**
Progression of phosphorylation from membrane to cytosol

Distribution of membrane proteins to response groups

- Proteins with maximal phosphorylation in early time points are enriched in membrane proteins.
- Proteins with maximal phosphorylation in later time points are enriched in soluble proteins.

# Functional properities of phosphoprotein nodes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinase</td>
<td>0.651</td>
<td>0.058</td>
<td>0.421</td>
<td>0.375</td>
<td>0.373</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Phosphatase</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.016</td>
<td>1</td>
<td>0.521</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Transport</td>
<td>0.433</td>
<td>1</td>
<td>1</td>
<td>0.442</td>
<td>0.698</td>
<td>0.297</td>
<td>0.360</td>
<td>0.634</td>
</tr>
<tr>
<td>Transcription</td>
<td>0.415</td>
<td>1</td>
<td>1</td>
<td>0.081</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Metabolic</td>
<td>0.346</td>
<td>0.222</td>
<td>0.614</td>
<td>0.551</td>
<td>0.751</td>
<td>0.531</td>
<td>0.706</td>
<td>1</td>
</tr>
</tbody>
</table>

p-Values of Fisher Exact Test for representation of biological functions in nodes of phosphorylation site network. Directionality from time information.

- Kinases are over-represented in single-multi nodes.
- Transcription factors and phosphatases are over-represented in single-single nodes.

GuangYou Duan; PhD-student
Mass-spectrometry based quantitative proteomics

Metabolic Labeling

Chemical Labeling
Proteins
Peptides

Synthetic Standard Peptides

Label-free Quantitation

# Mass-spectrometry Based Quantitative Proteomics

<table>
<thead>
<tr>
<th>Metabolic Labeling</th>
<th>Chemical Labeling</th>
<th>Synthetic Standard Peptides</th>
<th>Label-free Quantitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue</td>
<td>Proteins</td>
<td>Peptides</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectrum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproducible LC required for label-free quantitation

- Quantitative comparison can only be done between the same peptide sequence/charge state/modification.

- Averaging per protein only after standardization of peptide data.
Nano-LC-MS/MS Setup

Protein mixture

nano-HPLC  LTQ-Orbitrap mass spectrometer

C18-column (75μm), 8μm tip flow rate: 250 nL/min
Protein identification by mass spectrometry

Total ion chromatogram

MS full scan for Quantification

MS/MS fragment spectrum for Identification

AAVIGDTIGDPLK

Predicted protein sequences

Database Search → Peptide sequence → Peak List

Genome sequencing efforts
Protein inference problem in peptide-based proteomics

- Peptide sequences often cannot be assigned to a single protein only.
- Peptides "specific" to a protein: Proteotypic peptides.
- Often not proteins, but "protein groups" are identified by mass spectrometry.

Limitation to current "network" biology approaches

Proteome coverage for quantitation

- From all the proteins in the sample only a subset is being identified in LC-MS/MS experiments.
- Out of the identified proteins another subset is suitable for quantitation.
- Usually the more high abundant proteins are covered by identification and quantification.
- Fractionation is necessary for in-depth analysis.

Biggest challenge to current "network" biology approaches
Missing value problem

- Difficulty for statistical analysis and biological interpretation
- Stochastic component in classic dda LC-MS/MS leads to missing values.
- in different samples, different sets of proteins are identified / quantified.

Solutions:
- „Better“ instrumentation.
- Targeted analysis of protein subsets.
Application of Different Quantitative Methods

- **15N metabolic labeling**
  - Precision: +++
  - Coverage: ++
  - Dynamic range: ++

- **Label free approaches**
  - Precision: +
  - Coverage: +++
  - Dynamic range: ++

- **Chemical labeling**
  - Precision: ++
  - Coverage: ++
  - Dynamic range: ++

- **Targeted approaches using standard peptides**
  - Precision: +++
  - Coverage: +
  - Dynamic range: +++

- Different experimental setups require different choice of quantitative method.
- Depending on the biological question, the precision, coverage, and dynamic range must be considered.

cRacker as automated and standardized data analysis tool

- cRacker can process quantitative data from various quantitation softwares (MaxQuant, MSQuant, Progenesis, custom definitions).
- Processing of labeled and unlabeled data.
- Different possibilities for normalization and standadization.

\[ x_{\text{normalized}} = \frac{x_{\text{raw}}}{\sum x_{\text{raw}}} \]
\[ x_{\text{normalized}} = \frac{x_{\text{raw}}}{\sum x_{\text{raw}}} \cdot \frac{n_{\text{sample}}}{\max(n_{\text{sample.1}}, n_{\text{sample.2}}, \ldots, n_{\text{sample.n}})} \]
\[ x_{\text{scaled}} = \frac{x - \bar{x}}{\sigma} + \min(dataset) + \min(dataset) \cdot 10^{-6} \]
\[ x_{\text{scaled}} = \frac{(x_1, x_2, \ldots, x_n)}{\text{mean/median}(x_{\text{normalized}})} \]

Henrik Zauber; PhD-thesis


http://cracker.mpimp-golm.mpg.de
Towards a „network biology“

- Networks consist of nodes and connecting edges.
- Protein-protein interaction networks: undirected, direct physical interactions
- Protein-signaling networks: directed, not equal to PINs.

**Measured by:**
- AP-MS
- Y2H
- Cell biology

**Measurement ?**:
- Transient
- Often, edges need to be inferred

**Protein interaction networks are undirected**

**Protein-signaling networks are directed**

**Combined cellular network**