
Claude Nespoulous, Valérie Rofidal, Nicolas Sommerer, Sonia Hem, Michel Rossignol

To cite this version:

HAL Id: hal-01191388
https://hal.archives-ouvertes.fr/hal-01191388
Submitted on 1 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Phosphoproteomic analysis reveals major default phosphorylation sites outside long intrinsically disordered regions of *Arabidopsis* plasma membrane proteins

Claude Nespoulous\(^1\)*, Valérie Rofidal\(^1\), Nicolas Sommerer\(^2\), Sonia Hem\(^1\) and Michel Rossignol\(^1\)

**Background**

A large part of proteins in all eukaryotes, including plants, is predicted to contain intrinsically disordered regions (IDR), concerning long stretches of more than 30 residues, in a proportion depending on their subcellular localization [1]. Notably, by comparison to soluble proteins, transmembrane proteins are estimated to be richer in disordered regions [2] located at their cytoplasmic side, especially in the case of plasma membrane (PM) integral proteins [3]. In addition, direct assessment of IDR in published crystal structures for integral membrane proteins from various genomes and various subcellular origins showed that more than half of them actually display IDR [4]. At functional level, protein phosphorylation was suggested to occur predominantly in IDR [5]. In addition, in humans, recent proteome-wide data mining of curated information on post-translational modifications (PTM) confirmed that the frequency of phosphorylation is higher in predicted IDR and showed that this situation is mostly pronounced in the PM, where the enrichment of phosphosites within IDR reaches a factor of 2.7 [1]. Thus, as a general role of IDR in the adoption of structures favouring regulatory interactions...
is increasingly accepted [6], it could be speculated that integral membrane proteins use disordered regions for signalling and regulation, through various events such as reversible protein phosphorylation.

In plants, puzzling information is presently available on phosphorylation and IDR [7]. Overall, one third of protein sequences in Arabidopsis and rice genomes are predicted to constitute long IDR [1,8]. Actually, some soluble proteins were predicted with a high probability to have regulatory phosphosites within IDR. This concerns notably the dehydrin family of proteins [9] and the GRAS proteins whose disordered N-terminal domain constitutes the first functionally required unfolded domain in the plant kingdom [10,11]. By contrast, despite increasing functional evidence of the regulation of transport and signalling processes by phosphorylation, no specific information is available regarding membrane proteins, including the PM. In addition, several phosphorylation datasets have been generated from the Arabidopsis PM [12-18] and compiled into the PhosPhAt database (http://phosphat.mpimp-golm.mpg.de, [19]). But none of these analyses addressed the relationship between phosphorylation and IDR. Interestingly, studies in mice have shown that 86% of mouse brain phosphosites are located in predicted long IDR [20]. But presently in plants, only incomplete information is available and no effort was made to combine data about phosphoproteomics and disordered regions.

In this work, using a dedicated workflow, we identified novel phosphorylation sites in Arabidopsis PM vesicles. This set of phosphosites is shown to concern mainly integral PM proteins, mostly involved in transport and signal transduction. Thus, it allows to highlight original features regarding the location of phosphorylation sites in structured vs unstructured regions.

**Results and discussion**

Tryptic peptides from the enriched PM fraction were analyzed using a workflow designed to identify phosphorylated peptides and screen for novel phosphorylation sites (Additional file 1: Figure S1). For this purpose, we used a combination of peptide fractionation by Strong Anion eXchange (SAX) chromatography and phosphopeptide enrichment by TiO₂. This combination was previously shown to be able to identify specific subset of phosphorylation sites from PM transporters [16]. Tryptic peptides were then analyzed by LC-ESI MS/MS. Secondly, the resulting MS/MS data were queried against the Arabidopsis TAIR9 (http://www.arabidopsis.org/) database in the target-decoy mode in order to select peptides phosphorylated with 1% false discovery rate. The location of phosphorylated residues in these *bona fide* phosphopeptides was then assessed from their PTM score [21,22] and only unambiguous phosphorylation sites were selected. Finally, our dataset was compared with the PhosPhAt database and those sites that were not described previously were selected for further analysis. Overall, 411 phosphorylated peptides were characterized, resulting in the identification of 298 phosphoproteins and 559 unique phosphorylation sites. From this data set, the workflow allowed the characterization of 75 novel and unambiguous phosphorylation sites in 66 phosphopeptides corresponding to 52 protein accessions (Table 1, Additional file 2).

The distribution of modifications in terms of nature of the phosphorylated residues pSer/pThr/pTyr was found to be 75%/21%/4%, respectively. In addition, at the phosphorylation level, beside a majority of mono-phosphorylated species, multi-phosphorylated peptides accounted for one quarter of the total, of which 25% carried more than 2 modifications. The small size of the present phosphoproteome prevents definitive conclusions to be derived. However, the number of multi-phosphorylated peptides differed to some extent from previous Arabidopsis studies (often below 10% [14,15]) while the proportion of phosphotyrosines ranges between 0% and 4% as reported in PM [13-15] as well as in whole cell [23] studies. As similar cell culture conditions were used in previous Arabidopsis PM studies, the new features observed in our work do not arise from differences linked to the biological material. They should be thus attributed to our specific workflow and should be investigated further.

**The phosphoproteome displays features expected for PM proteins**

Plasma membrane vesicles were prepared from Arabidopsis cell suspension by differential centrifugation and phase partitioning between polyethylene glycol and dextran. Measurement of phosphohydrolase activities (Additional file 3: Figure S2) showed that the total Mg-dependent ATPase activity was over 95% sensitive to vanadate, a specific inhibitor of the plasma membrane H⁺-ATPase, and to a lesser extent to nitrate, with low azide-sensitive component. In addition, IDPase activity was below 5% of the ATPase activity. Collectively, this showed the prevalence of

**Table 1 Main features of proteins showing novel phosphosites**

<table>
<thead>
<tr>
<th>Features</th>
<th>Accessions</th>
<th>Phosphosites</th>
<th>TMD</th>
<th>PM prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>52</td>
<td>75</td>
<td>40</td>
<td>47</td>
</tr>
<tr>
<td>Transport</td>
<td>19</td>
<td>29</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Signalling</td>
<td>18</td>
<td>22</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>12</td>
<td>21</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Not assigned</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Features were computed from the resources described in main text. TMD: number of proteins with at least one trans-membrane domains (Aramemnon database); PM: number of proteins with a PM location (SUBA database). Proteins were classified in functional categories according to TAIR annotations and MapMan ontology.
phosphohydrolase activity of the Mg-dependent and vanadate-sensitive H⁺-ATPase type. This indicated that the membrane fraction was enriched in PM, with limited contamination by endomembranes, in agreement with results obtained using similar procedures for cell cultures or other samples from Arabidopsis [14,24].

According to the SUB-cellular location database of Arabidopsis proteins (SUBA, http://suba.plantenergy.uwa.edu.au/) and to TAIR annotations, 90% of the newly identified phosphoproteins were known or predicted to have a PM location (Table 1, “PM prediction” column). In addition, nearly 80% of accessions were predicted to display at least one trans-membrane domain (TMD) by the Aramemnon plant membrane protein database (http://aramemnon.botanik.uni-koeln.de/) (Table 1, “TMD” column).

Thus, both the above biochemical characterization of the membrane fraction and features of identified proteins indicate that the present subset of proteins corresponds mostly to genuine PM proteins, including a high proportion of integral proteins.

In order to get further information about the function of these proteins, we used both the Arabidopsis MapMan ontology (http://mapman.gabipd.org/web/guest/home) and TAIR annotations. The phosphoproteome identifies typical major PM functions. Indeed, two main categories emerged accounting for 72% of the dataset. The first category included protein transporters and the second proteins involved in signalling processes or protein phosphorylation. Each one of them consisted of more than one third of the total proteins (Table 1). The subset of transporters included both ion and small molecules transporters (e.g. phosphate transporters and auxin carriers) (Table 2). All of them were estimated to have a PM location and possessed between 6 and 15 TMD, at the exception of a magnesium transporter. This latter was predicted to display only 2 TMD and for which no previous information about a PM location was available.

A large part of signalling proteins corresponded to kinases from the Receptor-Like Kinases (RLK) superfamily. This family included notably members from the Leucine-Rich Repeat RLK (LRR-RLK) sub-family, most of them lacking yet a known role in a characterized process, with the exception of two alleles of brassinosteroid receptor BRI1. Proteins involved in protein phosphorylation covered various types of kinases, like Calcium-dependent Protein Kinases, together with one protein tyrosine phosphatase. Information about PM location was available for all proteins in these categories, and the presence of at least one TMD was predicted for 80% of them (Table 1). The privileged identification of these functional classes is in agreement with published proteomics and phosphoproteomics studies about the Arabidopsis PM [12,15,24].

The remaining 28% of the proteins consisted of 3 proteins of unknown function and 12 proteins belonging to various functional classes (Table 2). For these proteins, PM location was assumed at slightly lower rate (80%) and positive prediction of TMD concerned only nearly one half of them. Therefore, although this subset contained well-known PM proteins like the cellulose synthase A4, it is likely that part of these proteins could originate from other membrane systems, for example the golgin candidate 6, and/or have become adsorbed to PM vesicles during cell fractionation. In addition, these proteins accounted for a relatively small part of data by comparison to previous phosphoproteomics studies [12,15].

Collectively, the biochemical characterization of the membrane fraction, the function of the proteins identified and their features argue for a high content in true PM proteins in the present phosphoproteome, with prevalence for those involved in the exchange of solutes and information. In this view, the proposed strategy appears to generate information that complements available PM phosphoproteomics data. It thus enlarges by 15% the repertoire of experimentally determined phosphorylation sites in transporters and signalling proteins referenced at the PM in the PhosPhAt database (see below). Simultaneously, this dataset appears to be suitable to assess the localisation of phosphosites specifically in IDR from such proteins.

Phosphorylation is predicted to occur by default outside of long IDR

A number of algorithms were developed during the past decade for the prediction of IDR [25,26]. Regarding Arabidopsis, both pioneer estimations and recent analysis converge to a consensus of one third of proteins with a least one IDR of more than 30 residues [1,27]. In order to assess the extent of which the 75 novel phosphosites identified here could be located in such regions, we used the recent IUPforest-L predictor [28] for its high accuracy and efficiency (http://dmg.cs.rmit.edu.au/IUPforest/Eukaryota-L.php). Globally, from individual data from Table 2, the proportion of phosphorylation sites predicted to be located within long IDR amounted to 30% of the total number of phosphosites (Figure 1, black bars). This proportion held for proteins involved in transport or signalling, but was also true for proteins from other functional bins. In addition, at the phosphopeptide level, multi-phosphorylated peptides were found at a slightly lower rate (nearly 20%) within IDR than mono-phosphorylated ones. Thus, the present dataset of novel Arabidopsis PM phosphoproteins suggested a default prevalence of their phosphorylation sites within ordered regions, in comparison with current conclusions from combined analyses of long IDR predictions and
## Table 2 Identified phosphorylated proteins, peptides and novel sites

<table>
<thead>
<tr>
<th>Accession</th>
<th>Protein name</th>
<th>Peptide</th>
<th>Location</th>
<th>Sequence</th>
<th>Site</th>
<th>IDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G23080.1</td>
<td>Auxin efflux carrier family protein</td>
<td>VE[S][DV][S][LDGHD]{FLET}DAQ{GIDDGK}</td>
<td>S179</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G47670.1</td>
<td>transmembrane amino acid transporter family protein</td>
<td>V[S][T]PEIL[T]PSGQR</td>
<td>S19</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G48370.1</td>
<td>YELLOW STRIPE like 8</td>
<td>EEQEE[S]{V}EGIF{ESR}</td>
<td>S55</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G76430.1</td>
<td>phosphate transporter 1,9</td>
<td>SLEENE{D}{E}{Y}{S}({P}){AG}({G}){S}PANELLR</td>
<td>S522</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G01420.1</td>
<td>Auxin efflux carrier family protein</td>
<td>VE[S]{DV}{V}[S][LDGHD]{FLET}DAEIGNDGK</td>
<td>S179</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G01980.1</td>
<td>sodium proton exchanger, putative (NHX7) (SOS1)</td>
<td>QNTM{V}E[S]{D}E{D}{E}{G}{I}{V}{V}{R}</td>
<td>S1120</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G28070.1</td>
<td>ABC-2 type transporter family protein</td>
<td>QP{S}F{E}{D}{E}{S}{P}{E}{W}{E}{D}{T}{P}{V}{D}{L}{R}</td>
<td>S48</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G28120.1</td>
<td>Major facilitator superfamily protein</td>
<td>E[S]{P}{E}{S}{E}{L}{V}{P}{D}{S}{R}</td>
<td>S557</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G32830.1</td>
<td>phosphate transporter 1,5</td>
<td>EDEE{Q}{S}{G}{G}{D}{T}{V}{W}{F}{T}{V}{A}{N}{S}{G}R</td>
<td>T527</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G55320.1</td>
<td>putative subfamily B ABC-type transporter (AtMDR14)</td>
<td>E[S]{P}{E}{S}{E}{S}{E}{L}{V}{P}{D}{S}{R}</td>
<td>T774</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G23700.1</td>
<td>cation/H+ exchanger 17</td>
<td>NVT{E}{E}{S}{L}{V}{E}{D}{S}{E}{P}</td>
<td>S819</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G24120.1</td>
<td>YELLOW STRIPE like 1</td>
<td>EGGEE{D}{N}{N}{O}{L}{S}{Q}{E}{E}{E}{P}{D}{T}{E}{E}{M}{S}{G}R</td>
<td>T31</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G29900.1</td>
<td>autoinhibited Ca(2+)-ATPase 10</td>
<td>DVE{A}{G}{T}{S}{I}{F}{E}{Y}{E}{D}{S}{P}{E}{D}{I}{A}{S}{T}{K}</td>
<td>S22</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G01240.1</td>
<td>like AUXIN RESISTANT 1</td>
<td>QAE{S}{I}{V}{W}{S}{G}{E}{D}{E}{V}{A}{G}{R}</td>
<td>S14</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G24030.1</td>
<td>SLAC1 homologue 3</td>
<td>V{E}{D}{S}{I}{A}{A}{E}{E}{D}{I}{D}{I}{G}{N}{G}{F}{S}{M}{K}</td>
<td>S27</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G4350.1</td>
<td>phosphate transporter 1,1</td>
<td>V{L}{Q}{T}{D}{I}{E}{E}{E}{R}</td>
<td>T263</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G45380.1</td>
<td>sodium symporters;urea transmembrane transporters</td>
<td>SLEEL{G}{A}{E}{V}{S}{H}{D}{E}{K}</td>
<td>S520</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G64410.1</td>
<td>oligopeptide transporter 4</td>
<td>V{E}{A}{Y}{A}{S}{G}{D}{E}{V}{D}{V}{P}{A}{E}{E}{L}{R}</td>
<td>Y556</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G64560.1</td>
<td>magnesium transporter 9</td>
<td>E{I}{A}{G}{A}{Q}{N}{D}{D}{G}{T}{I}{G}{D}{E}{D}{E}{S}{P}{E}{F}R</td>
<td>T140</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G05150.1</td>
<td>Calcium-binding tetraetipptide family protein</td>
<td>AD{N}{N}{N}{N}{V}{D}{F}{S}{D}{A}{G}{W}{S}{J}R</td>
<td>S187</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G11330.1</td>
<td>S-locus lectin protein kinase family protein</td>
<td>S{G}{Q}{S}{L}{E}{L}{M}{N}{E}{W}{V}{S}{K}</td>
<td>S556</td>
<td>out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G53440.1</td>
<td>Leucine-rich repeat transmembrane protein kinase</td>
<td>LLL{D}{L}{T}{D}{V}{E}{I}{E}</td>
<td>T1030</td>
<td>in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Score</td>
<td>Spectrum</td>
<td>PhosphoPeptide</td>
<td>Modification</td>
<td>MW</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------</td>
<td>----</td>
</tr>
<tr>
<td>AT1G55610.1</td>
<td>BRI1 like</td>
<td>1</td>
<td>90</td>
<td>153</td>
<td>AD[T]EEDE[S]LDEFSLK</td>
<td>T1141 out</td>
</tr>
<tr>
<td>AT1G71860.1</td>
<td>Protein tyrosine phosphatase 1</td>
<td>46</td>
<td>118</td>
<td>119–136</td>
<td>AD[T]EEDE[S]LDEFSLK</td>
<td>S1146 out</td>
</tr>
<tr>
<td>AT3G13380.1</td>
<td>BRI1-like 3</td>
<td>47</td>
<td>106</td>
<td>119–136</td>
<td>TDLSAD</td>
<td>S</td>
</tr>
<tr>
<td>AT3G13530.1</td>
<td>Mitogen-activated protein kinase kinase kinase 7</td>
<td>55</td>
<td>109</td>
<td>481–510</td>
<td>VSEGKPNLEASTSMTPSNVQGE[S]VPAGDGK</td>
<td>S503 in</td>
</tr>
<tr>
<td>AT3G24660.1</td>
<td>Transmembrane kinase-like</td>
<td>2</td>
<td>54</td>
<td>109</td>
<td>K(T)(T)EDLEEGDEEDEIGEK</td>
<td>S334 in</td>
</tr>
<tr>
<td>AT3G25070.1</td>
<td>RPM1 interacting protein 4</td>
<td>59</td>
<td>30</td>
<td>37–61</td>
<td>IMNPNDPE[7]NDSQSAPPHPPSSR</td>
<td>Y45 in</td>
</tr>
<tr>
<td>AT3G28450.1</td>
<td>Leucine-rich repeat protein kinase family protein</td>
<td>2</td>
<td>69</td>
<td>128</td>
<td>[S]L</td>
<td>T</td>
</tr>
<tr>
<td>AT3G51740.1</td>
<td>Inflorescence meristem receptor-like kinase 2</td>
<td>1</td>
<td>65</td>
<td>155</td>
<td>[S]FSFDVGPNGGK</td>
<td>S31 out</td>
</tr>
<tr>
<td>AT4G23190.1</td>
<td>Cysteine-rich RLK (RECEPTOR-like protein kinase) 11</td>
<td>1</td>
<td>61</td>
<td>140</td>
<td>[S]GLTEVGVSGLAQR</td>
<td>S263 out</td>
</tr>
<tr>
<td>AT4G24630.1</td>
<td>DHHC-type zinc finger family protein</td>
<td>4</td>
<td>58</td>
<td>133</td>
<td>VEDDL[D]G</td>
<td>D</td>
</tr>
<tr>
<td>AT4G36180.1</td>
<td>Leucine-rich receptor-like protein kinase family protein</td>
<td>1</td>
<td>74</td>
<td>236</td>
<td>VGP</td>
<td>D</td>
</tr>
<tr>
<td>AT5G05160.1</td>
<td>Leucine-rich repeat protein kinase family protein</td>
<td>1</td>
<td>59</td>
<td>167</td>
<td>[S]E</td>
<td>W</td>
</tr>
<tr>
<td>AT5G19450.1</td>
<td>Calcium-dependent protein kinase 19</td>
<td>1</td>
<td>52</td>
<td>56</td>
<td>[S]FSFDVGPNGGK</td>
<td>S31 out</td>
</tr>
<tr>
<td>AT5G56890.1</td>
<td>Protein kinase superfamily protein</td>
<td>1</td>
<td>77</td>
<td>122</td>
<td>[S]LPHLIDNDVD[S]E[S]V[S]EAGDIGDR</td>
<td>S1049 in</td>
</tr>
<tr>
<td>AT1G10340.2</td>
<td>Ankyrin repeat family protein</td>
<td>4</td>
<td>72</td>
<td>31</td>
<td>[S]ADD</td>
<td>F</td>
</tr>
<tr>
<td>AT1G68720.1</td>
<td>tRNA arginine adenosine deaminase</td>
<td>72</td>
<td>119</td>
<td>109–1101</td>
<td>[S]F</td>
<td>E</td>
</tr>
<tr>
<td>AT3G09770.1</td>
<td>RING/U-box superfamily protein</td>
<td>68</td>
<td>67</td>
<td>290–317</td>
<td>Y</td>
<td>E</td>
</tr>
<tr>
<td>AT3G27530.1</td>
<td>Golgin candidate 6</td>
<td>74</td>
<td>271</td>
<td>897–914</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>AT3G28850.1</td>
<td>Glutaredoxin family protein</td>
<td>48</td>
<td>111</td>
<td>18–27</td>
<td>G</td>
<td>Y</td>
</tr>
<tr>
<td>AT4G18950.1</td>
<td>Integrin-linked protein kinase family</td>
<td>84</td>
<td>106</td>
<td>23–46</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>AT4G22670.1</td>
<td>HSP70-interacting protein 1</td>
<td>PM</td>
<td>55</td>
<td>59</td>
<td>(57–76)</td>
<td>SFVVEE[S]DDMDLTEEEKPK</td>
</tr>
<tr>
<td>AT5G07350.1</td>
<td>TUDOR-5N protein 1</td>
<td>PM</td>
<td>51</td>
<td>105</td>
<td>(965–984)</td>
<td>IGIWQYGDIE[S]DDEDTDGPAR</td>
</tr>
<tr>
<td>AT5G44030.1</td>
<td>cellulose synthase A4</td>
<td>PM</td>
<td>8</td>
<td>68</td>
<td>48</td>
<td>(81–100)</td>
</tr>
<tr>
<td>AT5G49720.1</td>
<td>glycosyl hydrolase 9A1</td>
<td>PM</td>
<td>1</td>
<td>55</td>
<td>152</td>
<td>(5–24)</td>
</tr>
<tr>
<td>AT5G62390.1</td>
<td>BCL-2-associated athanogene 7</td>
<td>PM</td>
<td>87</td>
<td>321</td>
<td>(429–446)</td>
<td>EIAEGVTIQVQLI[S]TTEE</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G45688.1</td>
<td>unknown protein</td>
<td>PM</td>
<td>1</td>
<td>64</td>
<td>190</td>
<td>(5–19)</td>
</tr>
<tr>
<td>AT3G27390.1</td>
<td>unknown protein</td>
<td>PM</td>
<td>5</td>
<td>55</td>
<td>219</td>
<td>(536–553)</td>
</tr>
<tr>
<td>AT5G64090.1</td>
<td>protein of unknown function</td>
<td>PM</td>
<td>1</td>
<td>50</td>
<td>131</td>
<td>(316–325)</td>
</tr>
</tbody>
</table>

PM SUBA, PM protein location according to the SUBA database; TMD, number of predicted trans-membrane domains according to the Aramemnon database; Mascot score, individual peptide Mascot; PTM score, post-translational modification score calculated using PhosCalc; in/out, inside/outside long IDR according to IUPforest-L predictor. Only peptides containing at least one novel and unambiguous phosphorylated residue are listed (square/normal brackets, unambiguous/ambiguous location according to the PTM score; see Materials and Methods).
experimental phosphosites identification in other organisms [5,20].

The origin of this unexpected finding is not clear and several causes can be envisaged. Firstly, the relatively small size of our dataset and the focus on novel sites might have introduced some bias. Accordingly, in order to average potential specific features of purified peptides, complementary information was searched by screening previously published data for other PM phosphorylated sites obtained from various methodological approaches.

For this purpose, the PhosPhAt database was searched for PM proteins that are classified in the bins 30 (signalling) and 34 (transport) of the Arabidopsis MapMan ontology. The location of the resulting set of phosphosites, including those identified as not novel in our study (187 and 160 in signalling and transport proteins, respectively) was then checked as above, using the IUPforest-L predictor (Additional file 4). Figure 1 (hatched bars) shows that less than 20% of sites were estimated to be located within long IDR, similarly for signalling proteins and transporters. Further combining this phosphoproteomics information with data from this work resulted in a set of 398 unambiguous phosphosites belonging to signalling and transport proteins, and showing similar distribution within IDR (20% and 22%, respectively).

Beside the nature of the dataset, another bias could arise from the prediction of IDR itself. Actually, it has been shown that predictors perform better with long IDR whom the boundary is currently established over 30 residues, than with shorter regions [29]. Large-scale statistics describing the relationship between IDR and phosphorylation rely indeed on IDR over a size of 30 residues, at the whole genome level as well as at the membrane level [1]. As the IUPforest-L predictor focuses on such IDR and ignore shorter disordered regions, we compared the results obtained for all proteins from our dataset of PM proteins with those from a consensus of 6 other predictors These predictors (DisEMBL, DorA, FoldIndex, GlobPlot2, RONN and IUPred) have complementary performances, not restricted to long IDR [30] and run simultaneously through the metaserver MeDor [31]. Figure 2 (black bars, panel d) shows that, on average, 30% of all phosphosites presented here could be proposed to be located in long IDR. However, another 28% of all phosphosites might be located in shorter regions (from 4 to 30 residues) also predicted as disordered (Figure 2, black bars, panels a, b and c). An equal proportion of about 10% was found for both intermediate sizes (11 to 20 and 21 to 30 residues) when only 5% of all phosphosites were predicted to be located in very short IDR (10 residues or less). In addition, this distribution held true for transport and signalling proteins. This situation was very predictor-dependent (Figure 2, grey bars), according to their own performance, with a proportion divergence around 40% for long IDRs and 15% for the 3 ranges described for short IDR for the total dataset. Moreover, just over 50% of phosphorylation sites were predicted to be located in regions, despite their size, simultaneously by at least 4 predictors as disordered (15% and 20% of short and long IDR, respectively, when took independently) and less than 8% by all the 6 predictors, (data not shown). Thus, together with results obtained above using IUPforest-L, this comparison might argue for a relatively limited distribution of our phosphosites set within long IDR (ca. 30%). Simultaneously, it pointed out on a possible role of short IDRs, although the low accuracy prediction [29] did not allow raising reliable conclusion. However, a potential contribution of short IDR, despite the low consensus of most of the predictions, might reach the proportion to over 50% of phosphosites of our dataset proposed as located in disordered regions regardless of the size.

Hence, both the extension of the present dataset to published data and the combination of several algorithms propose that a large part of phosphorylation events affecting the Arabidopsis PM proteins could occur by default outside long unordered regions of over 30 residues. As this notion is supported by a substantial number of phosphorylation sites (close to 400), the involvement of a sampling bias is unlikely. Similarly, as phosphorylation concerns cytosolic regions of PM proteins, the membranous nature of proteins is likely not responsible for the distribution observed. Thus, this supports the conclusion derived for the dataset obtained in this work and suggests that prevalent phosphorylation outside of long IDR might constitute a previously overlooked feature in Arabidopsis PM, specifically with transport and signalling proteins.
Examples of known regulatory phosphosites in *Arabidopsis*
are predicted to occur mainly outside of long IDR
Whereas the functional role of most PM phosphosites
above is presently unknown, for some of them a key role
was demonstrated in various types of regulations (not
listed in Table 2 which is restricted to novel sites). For in-
stance, the proton pumping ATPases of the AHA family,
constitute the major primary transporters that energize a
number of secondary active PM transporters. They are
well-known to be regulated by different phosphorylation
events, notably on the conserved penultimate Threonine,
which enables binding of regulatory 14-3-3 proteins and
results in ATPase activation [32]. But none of the 12 AHA
isoforms is predicted to display long IDR by IUPforest-L.
Similarly, the activity of the AMT1:1 ammonium trans-
porter was demonstrated to be regulated by phosphoryl-
ation of the conserved T460 in its C-terminal part [33].
But again no IDR is predicted in none of the 5 isoforms of
the AMT family. Except from activity regulation through
phosphorylation for different transporters, new evidence
recently showed that their trafficking is also controlled
by specific phosphorylation events. This is the case for
instance for the PIP2:1 isoform of aquaporins, where
phosphorylation of S283 was necessary for correct tar-
geting at the PM [34]. In this case also, none of the 8
isoforms of the PIP2-type of water channels is predicted
to display IDR. However, more complex situations can
be noticed for other transporters. In the case of phos-
phate transporters, the S514 residue of the PHT1;1 iso-
form is conserved in 7 out the 9 isoforms of the family
and its phosphorylation was shown to prevent the
PHT1:1 isoform from reaching its correct PM destin-
tation [35]. For this family, no long IDR is predicted in 6
isoforms, including PHT1:1. However the Serine residue
homologue to S514 is predicted to be located within a
long IDR in 2 out the 3 other isoforms. For signalling
proteins, that constitute the other main class character-
ized here, less data is available to assess the relationship
between regulatory phosphorylation events and disorder.
A notable exception concerns the BRII/BAK1 complex,
involving LRR-RLK that trans-phosphorylate each other
[36]. For this complex, the activation or inhibitory role
of respectively 12 and 5 phosphosites was demonstrated in
BRII and BAK1, several other sites being identified but
not characterized functionally. However, no IDR is pre-
dicted for BRII or for BAK1.
Thus, with some exceptions, the examples described above seem to illustrate well the functional aspect of the results obtained from the predictive analysis of a larger set. Thus the location of phosphorylation events affecting crucial functional features (such as activity, membrane targeting or protein interaction) does not appear associated with long disordered regions for the regulation of most of these transport and signalling proteins at Arabidopsis PM. The specific characteristics of such proteins did not allow extrapolating this conclusion to other proteins from other functional categories and with other physicochemical properties.

Conclusions

The phosphoproteomics data of plant plasma membrane proteins presented here, along with published datasets and available functional information, suggested a preferred topology of phosphorylation, at least regarding transport and signalling functions. This finding was unexpected according to the relationship of phosphorylation with long IDR usually reported, even after taking into account the questionable contribution of predicted short IDRs in our data set. Although the significance of this situation remains to be elucidated, two alternative or complementary hypotheses may be proposed. As about one half of Arabidopsis PM phosphosites would be located outside predicted disordered regions in proteins, it might be speculated that such observed phosphosites would have a more constitutive role rather than a regulatory role. The significance of such a role should be investigated in detail. Simultaneously, a part of regulatory phosphorylation events could be assumed to concern flexible, but of short length and more difficult to accurately predict, regions of proteins.

Methods

Samples

Arabidopsis (ecotype Col-0) suspension cells were grown in liquid Murashige and Skoog medium and ground in homogenisation buffer (100 mM Tris/HCl pH 8, 0.5 M sucrose, 10% glycerol (w/v), 0.6 % PVP (w/v), 10 mM EDTA, 10 mM EGTA, 10 mM ascorbic acid, 5 mM DTT, 1 mM PMSE, 1 μg/mL leupeptine) supplemented with, 1 mM sodium molydate, 1 mM orthovanadate, 50 mM sodium fluoride, 10 mM sodium pyrophosphate and 10 mM glycercophosphate to prevent protein dephosphorylation (as described elsewhere [16]). A crude membrane fraction was obtained by differential centrifugation (10 000 g max, 80 000 g max) and PM vesicles were extracted by two-phases partitioning using 6.4 % polyethylene glycol and dextran as in [37]. Purified vesicles were treated with 0.01% Brij58 to promote inside-out sidedness [38] and proteins (500 μg) were digested using trypsin (1/50, w/w; 37°C overnight), to obtain peptides from the cytoplasmic side.

Peptide purification

Strong Anion exXchange (SAX) microcolumns (packed in GELoader tips and equilibrated with 25 mM ammonium formate pH 7.5 and 30% acetonitrile) were used to fractionate PM peptides using increasing concentration of ammonium formate (6 steps from 25 mM to 1 M). Fractions were concentrated in vacuo to approximately 2 μL and diluted to 30 μL with 5% TFA in 80% acetonitrile prior to phosphopeptide selection. After loading on TiO2 microcolumns (prepared as in [39]) and washing with 1% TFA in 80% acetonitrile, phosphopeptides were eluted with 0.5% and 4.5% ammonium hydroxide.

Mass spectrometry and data analysis

Peptides were analyzed on an ion-trap Esquire HCT-plus mass spectrometer (Bruker) coupled to a ChipCube HPLC (Agilent). The chip contained both the pre-column and the column filled with the same stationary phase (Zorbax 300SB-C18; Agilent). Samples were first loaded onto the 4 mm enrichment pre-column at a flow rate of 4 μL/min using solvent A (0.1% formic acid). After pre-concentration, peptides were separated on the column (75 μm diameter, 150 mm length) at a flow rate of 0.3 μL/min using a 30 min linear gradient from 3% to 45% solvent B (0.1% formic acid, 90% acetonitrile) and eluted into the mass spectrometer. Raw MS data were processed using DataAnalysis and BioTools softwares (Bruker) to centroid spectra before querying the Arabidopsis TAIR9 database (http://www.arabidopsis.org; version pep_20090619) in the target-decoy mode using the Mascot search engine (Matrix Science; version 2.2.04). The following search parameters were used: up to one missed trypsin cleavage allowed, 1.2 Da mass tolerance for MS and 0.9 Da for MS/MS fragment ions; phosphorylation (ST) and (Y) as variable modifications. Under these conditions, for the dataset generated, a Mascot peptide score above 46 corresponded to 1% false discovery rate (FDR). For positive phosphopeptides, the probability based PTM score was calculated using the stand-alone software PhosCalc [22] including the algorithm developed by Olsen et al. [21], in order to assign individual phosphorylation sites. When different locations were computed, only those whose the score was higher than the maximum score minus five, were taken as unambiguous [21] and all others were rejected.

Additional files

Additional file 1: Figure S1. Work-flow for the identification of novel phosphorylation sites in Arabidopsis plasma membrane.

Additional file 2: Fragmentation data of novel and unambiguous phosphopeptides and sites.
Additional file 3: Figure S2. Phosphorydrolase activities of the membrane fraction.

Additional file 4: Phosphorylated proteins, peptides and sites from the bins 30 and 34 of the PhosPhAt data base. in/out, inside/outside long IDR according to IUPForest-L predictor. Only peptides containing at least one unambiguous phosphorylated residue are listed.

Abbreviations
IDR: Intrinsically disordered regions; PM: Plasma membrane; TMD: Transmembrane domain.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
CN performed both experimental steps upstream to mass spectrometry and bioinformatics analysis and wrote the paper. VR performed mass spectrometry experiments under the supervision of NS. SH analyzed phosphorylation data. MR designed the work and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Authors acknowledge the support of the Pole Proteome de Montpellier for mass spectrometry facilities and thank D. Centeno, E. Nicol and M. Tauzin for their support and N. Parisis and JB. Peltier for kindly rereading the manuscript.

Author details
1 UR1199 Laboratoire de Protéonomique Fonctionnelle, INRA, 34060 Montpellier cedex, France. 2UMR SPO, INRA, 34060 Montpellier cedex, France.

Received: 4 June 2012 Accepted: 23 October 2012 Published: 30 October 2012

References


