Pseudotargeted MS Method for the Sensitive Analysis of Protein Phosphorylation in Protein Complexes

Jiwen Lyu,†‡ Yan Wang,†‡ Jiawei Mao,†‡ Yating Yao,†‡ Shujuan Wang,§ Yong Zheng,§ and Mingliang Ye*†‡¶

†CAS Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic R&A Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences (CAS), Dalian 116023, China
‡University of Chinese Academy of Sciences, Beijing 100049, China
§State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

Supporting Information

ABSTRACT: In this study, we presented an enrichment-free approach for the sensitive analysis of protein phosphorylation in minute amounts of samples, such as purified protein complexes. This method takes advantage of the high sensitivity of parallel reaction monitoring (PRM). Specifically, low confident phosphopeptides identified from the data-dependent acquisition (DDA) data set were used to build a pseudotargeted list for PRM analysis to allow the identification of additional phosphopeptides with high confidence. The development of this targeted approach is very easy as the same sample and the same LC-system were used for the discovery and the targeted analysis phases. No sample fractionation or enrichment was required for the discovery phase which allowed this method to analyze minute amount of sample. We applied this pseudotargeted MS method to quantitatively examine phosphopeptides in affinity purified endogenous Shc1 protein complexes at four temporal stages of EGF signaling and identified 82 phospho-sites. To our knowledge, this is the highest number of phospho-sites identified from the protein complexes. This pseudotargeted MS method is highly sensitive in the identification of low abundance phosphopeptides and could be a powerful tool to study phosphorylation-regulated assembly of protein complex.

Proteins barely perform biological functions on their own, but interact with each other to execute an obligation together.1 Protein complex is a group of interacted proteins that work together to execute a specific biologic function.2 Protein phosphorylation plays a crucial role in regulating the assembling of protein complex.3,4 It is important to analyze the dynamical change of phosphorylation in the protein complex. Direct analysis of protein digest by shotgun proteomics for protein phosphorylation is very poor in sensitivity because the phosphopeptides are of low abundance in the sample and their ionization is seriously suppressed by the coexisted non-phosphopeptides.5 To reduce the interference of high abundant nonphosphopeptides, specific enrichment of phosphopeptides was often performed prior to LC-MS/MS analysis,6,7 which could identify and quantify over 10000 of phosphopeptides in a typical phosphoproteome analysis.8,9 Though the enrichment method performs very well for phosphoproteomics analysis, it might be not fitted to analyze trace amount of sample, especially in the protein complex sample which is often as low as a few microgram level. This is because huge sample loss may occur when enrichment was performed for trace amount of sample. Alternatively, the sample complexity can also be reduced by fractionation of the peptide sample prior to LC-MS/MS analysis, which allowed identification of low abundant peptides.10 To our knowledge, this multidimensional separation scheme was never applied to analyze protein complex probably because it also require large amount of sample.

Shotgun proteomics with mass spectrometer operated in data dependent acquisition (DDA) mode is a powerful tool for the discovery of new proteins but faces serious issues with reproducibility and sensitivity.11 More importantly, this data acquisition method is difficult to identify low abundant peptides as they have less chance to be delivered to MS2 for fragmentation.12 Targeted proteomics using either parallel reaction monitoring (PRM) or multiple reaction monitoring (MRM) has gained popular recent years due to its high sensitivity, high reproducibility, and high accuracy in quantification.13,14 It is of interest to develop targeted MS approach for the analysis of protein phosphorylation in protein complex. A classical targeted MS workflow includes two phases. First, in a pilot experiment for discovery, a large quantity of
sample is subjected to either 2D-LC-MS/MS analysis or phosphopeptide enrichment to maximize the identifications. Second, in a targeted analysis phase, the identifications of interest can be specifically monitored by 1D-LC-MS with MRM or PRM analysis to quantify across different samples. The sample amount for the pilot experiment was always over 100 μg for phosphorylation analysis.13–17 This amount is clearly too much for the development of targeted approach for the analysis of protein complex. In this study, we aim to develop a sensitive method for the identification and quantification of p-sites in the protein complex.

The recent developed quadrupole-Orbitrap offers specific trapping capacities to enhance the analysis of low abundance peptides.18,19 Compared with the MRM method in triple quadrupole MS, PRM in quadrupole-Orbitrap has higher resolution and selectivity.20 In addition to quantitative proteomics, PRM can also be applied in targeted peptide identification as it has all fragment ion information on the preselected precursor instead of giving only 3–5 preselected transitions in MRM.21 Taking the advantage of PRM’s high sensitivity and the feature of available full fragment ions in tandem spectra for peptide identification, we proposed a pseudotargeted MS method for the sensitive analysis of protein phosphorylation in protein complex. The development of this targeted approach is very easy as the same sample and the same LC-system with different acquisition modes, that is, DDA and PRM were used for the two phases. No sample fractionation or enrichment was required for the discovery phase which allowed this method to analyze minute amount of sample. This method was applied to analyze the phosphorylation in endogenous Shc1 complexes. With three runs of DDA in discovery phase and three runs of PRM in targeted analysis, totally about 3 μg was sufficient for the whole workflow and the identification of phospho-sites increased by around 50%. Take its advantage of and three runs of PRM in targeted analysis, totally about 3 μg was sufficient for the whole workflow and the identification of phospho-sites increased by around 50%. Take its advantage of and three runs of PRM in targeted analysis, totally about 3 μg was sufficient for the whole workflow and the identification of phospho-sites increased by around 50%. Take its advantage of and three runs of PRM in targeted analysis, totally about 3 μg was sufficient for the whole workflow and the identification of phospho-sites increased by around 50%.

**Enrichment of Phosphorylated Peptides by Ti-IMAC.** The sample pretreatment process was described in previous publications.6 Briefly, Ti-IMAC was added to the digest according to a ratio of 1:20 (peptides: beads). After 30 min incubation, the beads were washed by buffer 1 (50% ACN, 0.1% TFA, 200 mM NaCl) to remove the nonspecific peptides, followed with another wash step with buffer 2 (30% ACN, 0.1% TFA) to desalt. Finally, the phosphopeptides were eluted from beads by adding 10% ammonium hydroxide and dried down.

**Liquid Chromatography.** The dried samples were redissolved in 0.1% FA (1% FA only for Ti-IMAC enrichment sample) and loaded onto a 209 μm inner diameter 3 cm trap column (packed in-house with 5 μm, 120 Å, C-18 resins, from Sunchrom) using a flow rate of 5 μL/min of mobile phase A and washed for 10 min. H2O containing 0.1% formic acid was used as mobile phase A, while acetonitrile containing 0.1% formic acid was used as mobile phase B. Then the peptides were eluted from the trap column and the reversed-phase separation was accomplished using a 150 μm inner diameter 20 cm analytical column with a pulled tip (packed in-house with ReproSil-Pur C18-AQ 1.9 μm resin). The 130 min separation gradient was set as follow: the flow rate is 600 nL/min, 2% mobile phase B from 0 to 10 min, 7% B at 11 min, 27% B at 75 min, 45% B at 105 min, 90% B at 108 min, 90% B at 118 min, 2% B at 120 min, and 2% B at 130 min.

**Mass Spectrometry.** All data were acquired by Q-Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo, San Jose, CA) except the comparison with iDDA where Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer was used (refer to the Supporting Information). The DDA runs were conducted by a TopN method in which a high resolution (resolution of 60000 at 200 m/z) full MS acquisition was followed by 20 fast dd-MS2 acquisitions (resolution of 15000). The FTMS acquisition of full MS was set as followed: Automatic Gain Control (AGC) target, 3e6; Maximum injection time (Max IT), 20 ms; scan range, from 350 to 2000 m/z. The parameter of dd-MS2 acquisition was set as following: AGC target, 5e5; Max IT, 50 ms; isolation window, 1.6 m/z; normalized collision energy, 27%; Centriod mode. The ions that carrying charge lower than +1 and higher than +8 were excluded, and dynamic exclusion was set as 20 s.

The PRM runs were conducted under a Full MS and PRM tandem method with inclusion mode on. The isolation list for the identification of phosphopeptides was generated by sorting the potential phosphopeptides that identified by previous DDA.
the process of which is described in the next section. The MS was run on positive mode. For full MS scanning, the resolution was set at 60000, AGC target 3e6, Max.IT 20 ms and a profile mode was used for spectrum data. For the setting of MS2, which refers to PRM in this method, resolution was set at 30000, AGC target 5e5, Max.IT 503 ms, isolation window 1.6 m/z, NCE 27, and the profile mode was also used for spectrum data of MS2. For details of PRM quantification, please refer to Supporting Information.

Building of Pseudotargeted Library. The human proteome was obtained from UniProt (http://www.uniprot.org/) and included 20157 sequences of proteins. All of the *+.raw files that generated from DDA method were converted to *.MGF files by Thermo Proteome Discoverer (version 1.4). Mascot was used to search the *.MGF files against the database of human with parameter set as follow: precursor-ion mass tolerance, 10 ppm; fragment-ion mass tolerance, 0.05 Da; protease, trypsin with two missed cleavages. Variable modifications were set: oxidation on methionine (M, +15.9949 Da), phosphorylation on serine/threonine/tyrosine (S/T/Y, +79.9663 Da). After database searching, the proteins were filtered with <1% FDR. The proteins identified in each sample were used to generate a new focused database for reanalysis of the *.MGF files to identify potential phosphopeptides. For each matched phosphopeptides, the top-scored spectrum was extracted from the original searching DAT file via an in-house Java language package. Thus, the retention time, m/z, peptide sequence, peptide score and protein accessory of the phosphopeptides were available for the subsequent formation of PRM inclusion list. The phosphopeptides that did not belong to the protein complexes were removed and phosphopeptides that scored higher than 30 were also discarded as they were already confidently identified. The phosphopeptide identifications that conformed to the above criteria would be considered as targets for subsequent PRM acquisition. The isolation windows were set to ±3 min around the extracted retention time of the top-scored spectrum.

Data Analysis. All phosphorylation sites, which will be abbreviated as phospho-site or p-site, were identified by MaxQuant23 as it has a module to evaluate the p-site localization confidence. Mascot does not has this function, but its data is easy for us to extract the information on the retention time, m/z, peptide sequence that need for the generation of pseudotargeted library. The reviewed human proteome mentioned above was also used as a database in the identification of phospho-site. Trypsin was set as the specific enzyme and max miss cleavage was set at 2. Besides phosphorylation on S/T/Y, oxidation on methionine and acetylation on protein N-term were also set as variable modifications. For other items, the default settings were retained. The exported p-sites that localization probability less than 0.75 and score less than 40 were discarded. Quantitative data were processed with Skyline24 and MS2 ion chromatograms of interested peptides were used as surrogates for the quantification of proteins or phospho-sites. For details of label-free quantification (LFQ) and PRM quantification, please refer to Supporting Information. The raw data were uploaded onto JPOST Repository. The accession numbers are PXD008923 for ProteomeXchange and JPOST000385 for JPOST.

RESULTS AND DISCUSSION

Protein phosphorylation plays an important role in the dynamical assembling of protein complex. It is well-known that the EGF receptor tyrosine kinase (EGFR) is autophosphorylated upon EGF stimulation, which provide the binding sites for scaffold protein Shc1.27 Once Shc1 is associated with EGFR, it will be phosphorylated at multiple sites to provide docking sites for cytoplasmic targets. Therefore, Shc1 will recruit many proteins to form protein complexes upon EGF stimulation. Using EGF-dependent Shc1 complex as a test example, we first applied the conventional method to analyze its protein components and their phosphorylation after 2 min EGF stimulation. A special HeLa cell line expressing Flag tagged Shc1 was used for this study. The presence of Flag tagged Shc1 and the activation of EGFR were verified by Western blotting (Figure S1). When the Shc1 was pulled down by immunoprecipitation using anti-Flag antibody, its bona fide interaction proteins as well as a large number of nonspecific bound proteins were also pulled down.28 To distinguish these nonspecific bound proteins, the HeLa cells also expressing Flag tag, for example, Flag-GFP, would be an ideal control. However, such a cell line is not available in our lab. Instead, the wild type HeLa cells were used as the negative control in this study. After being stimulated with EGF for 2 min, the cells from these two cell lines were lysed and the Shc1 complex were immunoprecipitated under the same conditions. The obtained proteins were then digested and analyzed by LC-MS/MS in DDA mode as in the standard shotgun proteomics workflow. The identified proteins from these two samples were compared by label-free quantification using MaxLFQ26 and the proteins quantified with significant difference (p < 0.05) was obtained by t test (permutation-based FDR control, 1%) using Perseus.27 Only the proteins that exhibited significant high expression in the Flag-Shc1 HeLa cell line were considered as the potential components of the protein complexes (Figure S2). This resulted in the identification of 29 potential protein components. Since Shc1 recruits proteins to form protein complexes upon EGF stimulation, the bona fide complex components will increase in abundance in the pulled sample with EGF stimulation compared with the unstimulated one. Because of its high sensitivity, PRM was applied to monitor the dynamical change of the potential components in the Shc1 complex between the cells treated with EGF for 2 min or not. In addition to the 29 candidate proteins identified above, 7 additional proteins, PPP1R12A, PEAK1, AP2A2, PIK3CB, ERRF1, ASAP3, and PPP1CB, that were reported to be related components of the protein complexes (Figure S3a). These proteins were up-regulated over 50% (Figure S3a). These proteins together with Shc1, totally 34, were confirmed as the components of the EGF-dependent Shc1 complex. Among these 34 components, 20 were reported.22 The increases in abundance for five proteins, that is, PIK3C2B, ERBB3, ERBB2, SHC, and GRB2, after EGF stimulation were verified by Western Blotting (Figure S4). After the component proteins were determined, we then identified the phosphorylation sites on these proteins using DDA data acquired from the IP sample with 2 min EGF stimulation. We searched the DDA data against human proteome database using MaxQuant by setting variable phosphorylation modification on residues of Ser/Thr/Tyr. A total of 2166 unique peptides with score >40 were identified from 592 proteins, including 94 unique phosphopeptides. It was found 52 unique phosphopeptides were derived
from the 34 component proteins, which resulted in the identification of 42 p-sites with the localization probability higher than 0.75.

Using a conventional approach, we identified 42 p-sites from the Shc1 complex upon EGF stimulation for 2 min. We then proposed a pseudotargeted MS method as shown in Figure 1 to identify more p-sites from the protein complex. It is well-known that targeted analysis using PRM or MRM is more sensitive than conventional shotgun proteomics using DDA. Unlike MRM, PRM can provide full tandem spectra for peptide identification. PRM was usually used in targeted proteomics, but not in discovery proteomics, as the peptides to be analyzed must be set before LC-MS/MS analysis. Then how can we take the advantage of its high sensitivity to identify additional phosphopeptides with low abundance? The key issue is to build a library of potential phosphopeptides to be monitored. We reason that the potential list could be obtained from the DDA data. DDA is known to bias to identify high abundant peptides. It has poor sensitivity to identify low abundant phosphopeptides especially when the unphosphorylated peptides and phosphorylated peptides were loaded together for LC-MS/MS analysis. However, we believe that some low abundant phosphopeptides are also fragmented in DDA mode but failed to be identified using the strict filtering criteria. To discover these potential low abundant phosphopeptides, we used a focused database containing only the proteins identified in the sample for database searching of the DDA data. For the case of Shc1 complex sample, we first identified the proteins presented in the 2 min IP sample through searching the DDA data against human proteome database by MASCOT. These identified proteins were then used to construct a focused database, which was used to search the same DDA data to identify the potential phosphopeptides. No filtering criteria were applied for this step so that we can maximize the number of potential phosphopeptide identifications. From this step, we identified 1757 potential phosphopeptides. Among these, 222 phosphopeptides were derived from the component proteins of Shc1 complex. The high scored phosphopeptide identifications were removed as they were already confidently identified. After this step, 152 phosphopeptide identifications scored less than 30 were left. These identifications were of low confidence, but some true positive phosphopeptide identifications should be present in this list as the peptide mass and a few fragment ions were matched. Their masses, charges and retention times were used to build an inclusion list for PRM analysis. Considering the protein complex sample is not as complex as the proteome sample, relative wide retention time isolation window of 6 min was set to trap the potential phosphopeptides as much as possible. After the targeted analysis of these peptides by PRM, the resulting tandem spectra were searched against the human database. In this step, we used MaxQuant for database searching to identify phosphopeptides and localize p-sites. It should be noted two search engines, Mascot and MaxQuant, were used in this study. Mascot was used to generate the potential phosphopeptide library for PRM analysis because the retention time was more easy to be extracted by

Figure 1. Pseudotargeted strategy for the sensitive analysis of protein phosphorylation in protein complex. The low confident phosphopeptide identifications from DDA data were used to build the pseudotargeted library for PRM analysis, which enabled the identification of additional high confident phosphopeptides.
our in-house written Java script, while MaxQuant was used to identify phosphopeptides and phospho-sites in the final results because it has a module to evaluate the confidence of identified p-sites. After being filtered with the same strict criteria as in DDA data, 167 unique nonphosphopeptides and 43 unique phosphopeptides were confidently identified from the PRM data set. Because the presence of false matched phosphopeptides in the low-scored phosphopeptide target list, the nonphosphopeptides or phosphopeptides were also identified from proteins other than those presented in the protein complex. These identifications were removed and only the 37 unique phosphopeptides derived from Shc1 protein complex were kept. These phosphopeptides led to the mapping of 34 p-sites on 13 component proteins. Among the 34 p-sites, 19 were the same with the 42 p-sites identified by DDA. Therefore, 15 high confident p-sites were newly identified by this method. Thus, the total number of p-sites identified from 2 min stimulated sample was increased to 57. We named this method as the pseudotargeted MS method because we do not know if the targets we set are true positive target peptides or not. Among 152 potential phosphopeptides from the Shc1 complex that were targeted during PRM analysis, only 37 unique phosphopeptides were confidently identified from the protein complex after searching PRM spectra. Clearly most of them are not true positive targets. However, it should be noted that the scores for 49 peptides were improved (above the identical line) after PRM analysis (Figure 2a). One example is given in Figure 2b. A very poor MS/MS spectrum was generated in DDA, while a fragment rich spectrum was generated in PRM mode for the same phosphopeptide. The DDA spectrum cannot yield a confident phosphopeptide identification (Mascot score of 6) while the PRM spectrum yielded a high confident phosphopeptide identification with score of 34. Clearly the improved spectra quality in PRM enabled confident identification of extra p-sites in Shc1 protein complex.

In the above experiments, three replicate runs were performed and searched together for the analysis of the 2 min stimulated IP sample in both the DDA and PRM modes. Due to the high complexity of the proteome sample and the random sampling in DDA mode, running LC-MS/MS for multiple times is an effective way to improve the analysis coverage in shot-gun proteomics.30 We investigated if this is the case for the analysis of p-sites in protein complex. For easy comparison, the MS data for the three replicate runs were searched separately. As shown in Figure 3a, the three replicate DDA runs identified 35, 33, and 36 p-sites on the Shc1 complex, respectively. Combining the results of these three runs leaded to the identification of 43 p-sites (Figure 3a). The number did not increase significantly as most of these identifications were the same for different runs (Figure 3b). This is probably because the IP sample is much simpler than the common complex proteome sample and so that the mass spectrometer can reproducibly fragment these phosphopeptides. For the PRM analysis, the three replicate runs leaded to the identification of 30, 32, and 32 p-sites from the Shc1 protein complex (Figure 3c). Among them, 12, 12, and 14 p-sites were newly identified from the Shc1 protein complex compared with the p-sites already identified by DDA data. Combining these three PRM runs, the newly identified p-sites on Shc1 complex increased to 17. The increase for the three replicate PRM runs was even lower than the DDA data as the reproducibility for the targeted analysis was better (Figure 3d).

Figure 2. Pseudotargeted strategy improves the spectra quality. (a) Comparison of peptide matching scores in DDA and PRM for the same ions. Scores of 49 peptides were improved after PRM, locating above the identify line. (b) Matching score for the identification of a phosphopeptide, SPFGGPSAESVSSR, was 6 in DDA, while it was improved to 34 after PRM.

Clearly, replicate runs, either in DDA or PRM, does not contribute many new identifications. However, compared with the 42 sites identified by DDA, the PRM analysis operated in pseudotargeted strategy leaded to 15 new site identifications, which improved by 35.7%. Clearly the pseudotargeted strategy is able to identify low abundant phosphopeptides that cannot be achieved by conventional DDA method.

Inclusion list can also used in DDA mode, termed as iDDA, to enhance the detection sensitivity of the peptides of interest.30 We then compared the performance of iDDA with PRM using the 2 min stimulated IP sample. Filling time of 503 ms was set in PRM as it yielded maximum number of phosphopeptides during our initial optimization of PRM (data was not shown). For fair comparison, filling time of 503 ms was also set in iDDA rather than 50 ms in conventional DDA we mentioned above. Basically, the comparison experiments...
were performed in parallel with identical conditions, that is, the same sample, the same inclusion list, and the same machine (QE rather than QE HF was used here) with the same parameter setting. It can be seen from Figure S5a that both iDDA and PRM yielded much more p-site identifications than conventional DDA did. However, the p-sites identified by PRM were 37.5% more than those identified by iDDA. Clearly PRM has higher sensitivity. This could be partly attributed to their different trigger mechanism. In iDDA, MS2 were triggered only when the presence of the peaks in the full MS corresponding to m/z values set in the inclusion list. While in PRM, MS2 were triggered by preset retention times and m/z values independent of MS scan. The low abundance phosphopeptides may not yield strong enough peaks in MS1 to trigger MS2 in iDDA; however, these MS2 spectra can still be available in PRM mode. This explained why more MS2 spectra from the protein complex were collected in PRM mode compared with in iDDA mode (Figure S5b).

This PRM based pseudotargeted method was further applied to analyze the Shc1 complexes with other EGF stimulation times. Obvious increases in the number of the identified p-sites were observed in all the samples as shown in Table 1. Especially 21 more p-sites were identified by PRM in the 20 min sample, which accounted for 50% of the p-sites identified by DDA. From this study, we identified 10, 57, 63, and 63 p-sites in Shc1 complexes at 4 states of EGF signaling pathway (rest stage/0 min, early stage/2 min, medium stage/5 min, late stage/20 min), respectively. Accumulatively, a total of 82 unique p-sites were identified in this study. The identified p-sites were listed in Table S2. It was found that over 96% of these p-sites were included in PhosphoSitePlus, a database of experimentally observed post-translational modifications. To our knowledge, this is the highest number of p-sites identified from the EGF-dependent Shc1 complexes. In a previous work, 22 p-sites were identified on the Shc1 complexes. With the benefits of high sensitivity and high resolution of PRM, the identification of p-sites in the protein complex was dramatically enhanced by the pseudotargeted strategy.

In conventional phosphoproteomics analysis workflow, the phosphopeptides were specifically enriched from protein digest prior to LC-MS/MS run. Because the interferences from the nonphosphopeptides were almost eliminated, the phosphoproteomics coverage could be dramatically improved. We also tested the performance of this approach for the analysis of the p-sites in protein complex. The digest of the 20 min EGF stimulated Shc1 complex IP sample, which has been identified most p-sites in both acquisition modes, was subjected to Ti-IMAC enrichment followed by LC-MS/MS analysis. It was found only five p-sites on the Shc1 complex were identified, which is much fewer than the identification in sample without Ti-IMAC treatment (see Table 1). This is not difficult to understand. The pulled down complex is about a few μg, which is much fewer than the common initial amount for Ti-IMAC enrichment of proteome sample (typically >100 μg). The

Table 1. P-Sites Identified in the Shc1 Protein Complexes by DDA and PRM

<table>
<thead>
<tr>
<th>EGF stimulation time</th>
<th>0 min</th>
<th>2 min</th>
<th>5 min</th>
<th>20 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDA</td>
<td>6</td>
<td>42</td>
<td>46</td>
<td>42</td>
</tr>
<tr>
<td>PRM</td>
<td>+4</td>
<td>+15</td>
<td>+17</td>
<td>+21</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>57</td>
<td>63</td>
<td>63</td>
</tr>
</tbody>
</table>

Figure 3. Multiple runs of DDA or PRM analysis do not significantly improve the coverage. The numbers of p-sites identified by the three replicate (a) DDA runs and their combination, (c) PRM runs and their combination; The overlap of p-site identifications for the three replicate (b) DDA and (d) PRM runs. The Shc1 complex immunoprecipitated from cells stimulated with EGF for 2 min was used as the sample for all the analysis. The p-sites shown were all identified from the Shc1 complex. The numbers of p-sites for the combined results were slightly different with those in the Table 1 because the MS data for each run were searched separately here.
enrichment step generated huge sample loss due to the minute amount of sample, which resulted in poor identification. Alternatively, the method presented in this study without enrichment yielded better results.

Due to its advantage of accurate quantification, the PRM based method enabled comprehensive mapping the dynamic change of the shc-1 complex upon EGF stimulation (Figure 4). After EGF stimulation, a total of 33 proteins were up-regulated after EGF stimulating, among which 12/15/6 proteins achieved peak point at 2 min/5 min/20 min, respectively (Figure S3b). Some proteins were down-regulated in later stages, as the intensity decreased by 50% comparing to the peak value but still maintained at a remarkable high level. Similarly, dynamical change in protein phosphorylation on the protein complexes was also observed in different stages of EGF stimulation. The numbers of p-sites identified on each protein were labeled on Figure 4. With the assembling of protein complex and signal transduction, the total number of p-sites reached maximum at late stage (20 min after EGF stimulation). About 42.7% p-sites were identified in all EGF stimulated states (2 min/5 min/20 min) (Figure S6), suggesting these sites started to be activated immediately after the cells were stimulated and most of p-sites maintained activated from early stage to late stage. The peptides that carrying the p-sites were monitored by PRM and the dynamic change of 45 p-sites across the stimulation was observed (Table S1 and Figure S7a). Changes in levels of phosphopeptides were further normalized to changes in protein expression to derive changing tendency in occupancy of phosphorylation sites (Figure S7b). The normalized fold-change indicates the change of phospho-site occupancy. It was found the occupancy for many sites altered during the EGF stimulation, suggesting phosphorylation-mediated dynamic regulation of protein complex. For example, the quantitative tendencies of Y1172 and Y1197 on EGFR were observed to reach peaks after 5 min stimulation, which is similar to the result of previous work that the two p-sites reached peaks at early stage and then decreased. After normalized by protein change, the occupancy of these two sites started to increase at 2 min and were down-regulated at 5 min. The occupancies that seemed high at 20 min may be the result of down-regulation in protein level. The dynamic occupancy changes on these two sites suggested that the sites started to be activated immediately after stimulation and did not increase as the same proportion as the increasing of protein level. Bring together, the above results indicated that the PRM based method could be a powerful tool to reveal the process for the dynamical assembly of protein complex.

CONCLUSION

In this study, we presented a pseudotargeted MS method to identify low abundance phosphorylation in minute amount of sample. The development of this targeted approach is very easy as the same sample and the same LC-system were used for the two phases. No sample fractionation or enrichment was required for the discovery phase which allowed this method to analyze minute amount of sample. We have demonstrated that this method has higher sensitivity to identify phosphor-ylation sites on endogenous Shc1 protein complexes. PRM is typically used in targeted proteomics to quantify the peptides of interest, while the pseudotargeted MS method presented in this study allowed PRM to identify new peptides. In this new method, low confident peptides identified from the DDA data set are used to build a pseudotargeted library for PRM analysis which enabled the identification of new high confident peptides. Thus this strategy is not limited to analyze protein

Figure 4. PRM-based method enabled comprehensive mapping the dynamic change of the Shc-1 complex upon EGF stimulation. The binding partners of Shc1 at rest phase were labeled in blue. The proteins recruited after EGF stimulation was shown in orange. The proteins down-regulated over 50% compared with their peak values were labeled in white. The number in the circle indicates the number of p-sites identified from that protein.
phosphorylation. It is also applicable to analyze other low abundant peptides in a sample with trace amount.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b00749.

Supplementary results mentioned in the main text and Figures S1–S7 (PDF).

Table S1: The analysis of the Shc1 complexes by PRM (XLSX).

Table S2: The p-sites identified in this study (XLSX).

AUTHOR INFORMATION

Corresponding Author
*Phone: +86-411-84379610. Fax: +86-411-84379620. E-mail: mingliang@dicp.ac.cn.

ORCID

Mingliang Ye: 0000-0002-5872-9326

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported, in part, by funds from the China State Key Basic Research Program Grants (2016YFA0501402, 2017YFA0505004), the National Natural Science Foundation of China (21535008, 81600046, 91753105). MY is a recipient of the National Science Fund of China for Distinguished Young Scholars (2017YFA0505004), the National Natural Science Foundation of China (20141140116, 81500046, 91753105). MY is a recipient of the National Science Fund of China for Distinguished Young Scholars (2017YFA0505004). The authors declare no competing financial interest.

REFERENCES