Identification of Extracellular Signal-regulated Kinase 1 (ERK1) Direct Substrates using Stable Isotope Labeled Kinase Assay-Linked Phosphoproteomics*

Liang Xue‡, Pengcheng Wang§, Pianpian Cao¶, Jian-kang Zhu‡§, and W. Andy Tao‡∥‡‡§§

Kinase mediated phosphorylation signaling is extensively involved in cellular functions and human diseases, and unraveling phosphorylation networks requires the identification of substrates targeted by kinases, which has remained challenging. We report here a novel proteomic strategy to identify the specificity and direct substrates of kinases by coupling phosphoproteomics with a sensitive stable isotope labeled kinase reaction. A whole cell extract was moderately dephosphorylated and subjected to in vitro kinase reaction under the condition in which 18O-ATP is the phosphate donor. The phosphorylated proteins are then isolated and identified by mass spectrometry, in which the heavy phosphate (±85.979 Da) labeled phosphopeptides reveal the kinase specificity. The in vitro phosphorylated proteins with heavy phosphates are further overlapped with in vivo kinase-dependent phosphoproteins for the identification of direct substrates with high confidence. The strategy allowed us to identify 46 phosphorylation sites on 38 direct substrates of extracellular signal-regulated kinase 1, including multiple known substrates and novel substrates, highlighting the ability of this high throughput method for direct kinase substrate screening. Molecular & Cellular Proteomics 13: 10.1074/mcp.O114.038588, 3199–3210, 2014.

Protein phosphorylation regulates almost all aspects of cell life, such as cell cycle, migration, and apoptosis (1), and deregulation of protein phosphorylation is one of the most frequent causes or consequences of human diseases including cancers, diabetes, and immune disorders (2). Up till now, however, known substrates are far from saturation for the majority of protein kinases (3); thus, mapping comprehensive kinase-substrate relationships is essential to understanding biological mechanisms and uncovering new drug targets (4).

Accompanied with advances of high-speed and high-resolution mass spectrometry, the technique of kinase substrate screening using proteomic strategy is quickly evolving (5–7). Mass spectrometry has been extensively used for kinase-substrate interaction mapping (8) and global phosphorylation profiling (9). Although thousands of phosphorylation sites have been detected, complex phosphorylation cascade and crosstalk between pathways make it difficult for large-scale phosphoproteomics to reveal direct relationships between protein kinases and their substrates (10, 11). Extensive statistics, bioinformatics, and downstream biochemical assays are mandatory for the substrate verification (12, 13). Another strategy uses purified, active kinases to phosphorylate cell extracts in vitro, followed by mass spectrometric analysis to identify phosphoproteins. This approach inevitably faces the major challenge of separating real sites phosphorylated by target kinase and the phosphorylation triggered by endogenous kinases from cell lysates (14). Analog-sensitive kinase allele (15) overcomes the issue by utilizing the engineered kinase that can exclusively take a bulky-ATP analog under the reaction condition. Analog-sensitive kinase allele has been coupled with γ-thiophosphate analog ATP to facilitate the mass spectrometric analysis (16–18).

We have introduced kinase assay-linked phosphoproteomics (KALIP)1 to link the in vitro substrate identification and...

From the ‡Departments of Biochemistry, §Horticulture and Landscape Architecture, ¶Mathematics, ||Medicinal Chemistry and Molecular Pharmacology, **Chemistry, and ‡‡Purdue University Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907
Received February 12, 2014, and in revised form, July 7, 2014
Published, MCP Papers in Press, July 14, 2014, DOI 10.1074/mcp.O114.038588

Author contributions: L.X., J.Z., and W.A.T. designed research; L.X. and P.W. performed research; L.X., P.C., and W.A.T. analyzed data; L.X. and W.A.T. wrote the paper.

1 The abbreviations used are: BAG3, BAG family molecular chaperone regulator 3; KALIP, kinase assay-linked phosphoproteomics; regulator 3; BORG4, Cdc42 effector protein 4; EPS15, epidermal growth factor receptor pathway substrate 15; ERK1, extracellular signal-regulated kinase 1; FASP, filter aided aided proteome preparation; FDR, false discovery rate; FSBA, adenosine-5′-(4-fluorosulfonylbenzoyl) hydrochloride; GO, gene ontology; IPA, ingenuity pathway analysis; MAPK, mitogen-activated protein kinase; PolyMAC, polymer-based metal-ion affinity capture; proKALIP, protein kinase assay linked with phosphoproteomics; QIKS, quantitative identification of kinase substrates; SILAC, stable isotope labeling by amino acids in cell culture; siKALIP, stable isotope labeling kinase assay-linked phosphoproteomics.
physiological phosphorylation events together in a high throughput manner (19, 20). The strategy, however, has only been applied to identify direct substrates of tyrosine kinases. In this study, we expanded the application of KALIP to serine/threonine kinases by introducing a quantitative strategy termed Stable Isotope Labeled Kinase Assay-Linked Phosphoproteomics (siKALIP). The method was used to identify direct substrates of extracellular signal-regulated kinase 1 (ERK1), a serine/threonine kinase acting as an essential component of the Mitogen-activated protein kinase (MAPK) signal transduction pathway (21). A defect in the MAP/ERK pathway causes uncontrolled growth, which likely leads to cancer (22) and other diseases (23–25). ERK1 can be activated by growth factors such as platelet-derived growth factor (PDGF), epidermal growth factor (EGF), and nerve growth factor (NGF) (26). Upon stimulation, ERK1 phosphorylates hundreds of substrates in various cellular compartments including cytoplasm, nucleus, and membrane (27). Among 38 ERK1 direct substrates identified by siKALIP, more than one third are previously discovered by classical molecular biology approaches, highlighting high specificity and sensitivity of the strategy. The results also support the hypothesis that ERK1 plays complex roles in multiple pathways that are essential for the cell growth regulation.

**EXPERIMENTAL PROCEDURES**

**Mammalian Cell Culture**—HEK293 cells (ATCC) were maintained in Dulbecco’s Modified Eagle Medium (DMEM) (Sigma) supplemented with 10% heat inactivated FBS, 100 μg/ml streptomycin, and 100 IU/ml penicillin in 5% CO2 at 37 °C. Human DG-75 B lymphoma cells (ATCC) were grown in RPMI 1640 media (Sigma) supplemented with 10% heat inactivated FBS, 1 mM sodium pyruvate, 100 μg/ml streptomycin, 100 IU/ml penicillin, and 0.05 mM 2-mercaptoethanol in 5% CO2 at 37 °C. The cells were washed with PBS, trypsinized, collected, and frozen at −80 °C for further use.

**Plant Tissue Culture**—Seedlings of *A. thaliana* were grown in 40 ml of half-strength Murashige and Skoog medium at 22 °C in continuous light on a rotary shaker set at 100 rpm. Twelve-day-old seedlings were collected and frozen at −80 °C for further use.

**In Vitro Kinase Reaction in siKALIP**—Mammalian cells were lysed by sonication in lysis buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, and 1% Nonidet P-40 on ice. For plant cells, total protein was extracted from 2 g of seedlings by grinding in 2 ml of extraction buffer containing 100 mM Tris-HCl (pH 7.5), 250 mM NaCl, and 5 mM EDTA. The cell debris was cleared by centrifugation at 16,000 × g for 10 min. The supernatant containing 400 μg soluble proteins was collected. The lysate volume was adjusted to 200 μl using phosphatase buffer (Roche). 10U of rPAkd alkaline phosphatase (Roche) was added and incubated at 37 °C for 3 h. The phosphatase was deactivated by heating at 75 °C for 5 min. To inhibit endogenous kinases, the sample was incubated with 1 mM 5’-(4-fluorosulfonylbenzoyl)adenosine (FSBA) with 10% DMSO in Tris-HCl, pH 7.5 at 30 °C for 1 h. Excess FSBA is removed by Vivaco filtration units (30 kDa cutoff). Samples in the filters were then incubated in buffer containing 300 ng ERK1 (Sigma), 5 mM MgCl2, and 1 mM γ[32P]ATP (Cambridge Isotope Laboratory, Andover, MA) at 30 °C for 1 h. The reaction was stopped by 8 M urea with 5 mM dithiothreitol. Filter aided proteome preparation (FASP, San Diego, CA) digestion was performed according to the manufacturers’ specifications (Expedeon).

**In Vitro Kinase Reaction by Autoradiography**—The EPS15 and BAG3 substrates were isolated on beads by specific antibody (Anti-EPS15 from Cell Signaling, Beverly, MA and Anti-BAG3 from Protein-Tech, Chicago, IL). The BORG4 recombinant proteins were purchased from Abnova. Substrates were incubated in Tris-HCl, pH 7.5 buffer containing 1 μg ERK1 (Sigma), 5 mM MgCl2, and 25 μM cold ATP, 2.5 μCi [γ-32P]ATP at 30 °C for 1 h. The reaction was quenched by boiling the sample with 4X electrophoresis sample buffer (Invitrogen, Carlsbad, CA). Phosphorylation signal was detected by phospho-imager (GE Healthcare, Pittsburgh, PA).

**Phosphopeptide Enrichment**—Tryptic peptides were first desalted using a Sep-pak C18 column (Waters, Milford, MA) and dried. Next, the peptide mixture was resuspended in 100 μl of loading buffer (100 mM glycine buffer, 1% trifluoroacetic acid, and 50% acetonitrile) to which 5 nMol of the PolyMAC-Ti (Tymora Analytical, IN) reagent was added (28). The mixture was then incubated for 5 min. 200 μl of 300 mM HEPES, pH 7.7, was added to the mixture to achieve a final pH of 6.3. The solution was incubated with magnetic hydrazide beads to capture the PolyMAC-Ti dendrimers. The column was gently agitated for 10 min and then centrifuged at 2300 × g for 30 s to collect the unbound flow-through. The beads were washed once with 200 μl loading buffer, twice with a mixture of 100 mM acetic acid, 1% trifluoroacetic acid, and 80% acetonitrile, and once with water. The phosphopeptides were eluted from dendrimers by incubating the beads twice with 100 μl of 400 mM ammonium hydroxide for 5 min. The eluates were collected and dried under vacuum.

**Mass Spectrometric Data Acquisition**—Peptide samples were dissolved in 8 μl of 0.1% formic acid and injected into an Eksigent NanoLC Ultra 2D HPLC system. The reverse phase chromatography was performed using an in-house C18 capillary column packed with 5 μm C18 Magic beads resin (Michrom; 75 μm i.d. and 12 cm bed length). The mobile phase buffer consists of 0.1% formic acid in ultra-pure water with an eluting buffer of 0.1% formic acid (Buffer A) in 100% CH3CN (Buffer B) run over a linear gradient (2–35% Buffer B, 90 min) with a flow rate of 300 nl/min. The electrospray ionization emitter tip was generated on the prepacked column with a laser puller (Model P-2000, Sutter Instrument Co.). The Eksigent Ultra2D HPLC system was coupled online with a high-resolution hybrid dual-cell linear ion trap Orbitrap mass spectrometer (LTQ-Orbitrap Velos; Thermo Scientific). The mass spectrometer was operated in the data-dependent mode in which a full-scan MS from m/z 300–1700 with the resolution of 60,000 at m/z 400 was followed by 20 MS/MS scans of the most abundant ions using collision-induced dissociation (CID). Ions with the charge state of +1 were excluded. The dynamic exclusion time was set to 60 s after two fragmentations.

**Database Search and Quantitation**—The LTQ-Orbitrap raw files were searched directly against the Homo sapiens database with no redundant entries (93,289 entries; human International Protein Index (IPI) v.3.83) using the SEQUEST algorithm on Proteome Discoverer (Version 1.3; Thermo Fisher). Peptide precursor mass tolerance was set to 10 ppm, and MS/MS tolerance was set to 0.8 Da. Search criteria included a static modification of +57.0214 Da on cysteine residues, a dynamic modification of +15.9949 Da on oxidized methionine, a dynamic modification of +79.996 Da on normal phosphorylated serine, threonine, and tyrosine residues, and a dynamic modification of +85.996 Da on heavy phosphorylated serine, threonine, and tyrosine residues. Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on peptides analyzed by the sequence database. False discovery rates (FDR) were set to 1% for each analysis. Proteome Discoverer generated a reverse “decoy” database from the same protein database, and any peptide passing the initial filtering parameters from this decoy database was defined as a false positive. The minimum cross-correlation factor (Xcorr) filter was re-adjusted for each charge state separately in order
to optimally meet the predetermined 1% FDR based on the number of random false-positives matched with the reversed “decoy” database. Thus, each dataset had its own passing parameters. The number of unique phosphopeptides and nonphosphopeptides were then manually counted and compared. Phosphorylation site localization from CID mass spectra was determined by PhosphoRS scores (29). For phosphopeptides with ambiguous phosphorylation sites, only one phosphorylation site with the highest score was selected for further data interpretation.

Data Analysis—To find the ERK1 phospho-motif analysis, Motif-X (http://motif-x.med.harvard.edu) was used for predicting the specificity of kinases according to identified phosphorylation sites. Parameters were set to peptide length $\leq 13$, occurrence $\geq 10$, and significance $p$ value less than 0.000001. For the ERK1 binding site analysis, Scansite 2.0 (http://scansite.mit.edu) was used to search Uniprot protein sequences for high-stringency ERK1/2 binding sequences. The heavy phosphorylated proteins in vitro were submitted to Ingenuity Pathway Analysis (IPA) (Ingenuity Systems) for the gene ontology (GO) annotation.

Immunoprecipitation and Western Blotting Experiments—Cells were collected and lysed by sonication in lysis buffer containing protease and phosphatase inhibitors (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 1% Nonidet P-40, and 1x Mini Complete protease inhibitor mixture (Roche)). Samples were cleared of debris and normalized based on the protein concentration. Then 1 mg of lysate was pre-incubated with 20 $\mu$g Protein A/G agarose beads (Thermo) for 20 min at 4 °C to remove nonspecific bindings, before further incubation with 10 $\mu$g of antibodies for 4 h at 4 °C. Anti-EPS15 rabbit polyclonal antibody from Cell Signaling and anti-BAG3 rabbit polyclonal antibody from Proteintech were used. The samples were then incubated with 20 $\mu$g of Protein A/G agarose beads again for capturing for 4 h at 4 °C. The beads were washed and bound proteins were eluted by boiling the beads in the SDS loading buffer with 50 mM DTT for 5 min. The eluents were separated on a 12% SDS-polyacrylamide gel and transferred onto a PVDF membrane. The membranes were probed using antibodies against proteins of interest.

RESULTS

The siKALIP Strategy for Direct Kinase Substrate Identification—In contrast to tyrosine phosphorylation, it is estimated that serine/threonine phosphorylation consists of over 99% of all phosphorylation (30). In a typical whole cell extract, thousands of proteins can be phosphorylated on serine/threonine residues (31) and present enormously large background in any phosphorylation analysis. It poses great challenges to use whole cell extracts for in vitro kinase assay. To address the issue, we devised a new strategy, siKALIP, by incorporating $\gamma$-18O-phosphate ATP and moderate dephosphorylation in the kinase assay to identify direct kinase substrates in high throughput (Fig. 1). In general, proteins are extracted from the cells of interest. A phosphatase is added to partially remove phosphate groups from endogenous phosphoproteins to generate a pool of the candidate proteome for the following in vitro kinase reaction. Pulsed heating is applied to quench the exogenous phosphatase activity. A generic kinase inhibitor such as FSBA is added to deactivate endogenous kinase activities and excess FSBA was subsequently removed using filtration. The protein solution is then equally divided prior to the addition of the kinase along with $\gamma$-[18O] ATP (sample) or $\gamma$-[18O] ATP only (control). After the in vitro kinase reaction, proteins are digested and phosphopeptides are enriched by Polymer-based Metal ion Affinity Capture (PolyMAC) (28), fol-
Identification of ERK1 Direct Substrates by siKALIP

Fig. 2. A, Dephosphorylation and $^{18}$O-labeled ATP in siKALIP. The cell lysates are dephosphorylated and then incubated with $[^{18}\text{O}_{4}]$-ATP under the reaction condition. Samples at the basal level, and after the kinase-treated, along with the control are compared. Red and green phosphates indicate $[^{18}\text{O}_{3}]$ and $[^{18}\text{O}_{4}]$, respectively. The peaks of phosphopeptides are labeled with different colors (yellow, blue, and purple). Yellow: The endogenous phosphorylation is not completely removed, so the heavy and light isotopic peaks exist. Blue: The endogenous phosphorylation is removed, and only the heavy phosphorylation exists after kinase assay. Purple: endogenous phosphorylation sites not detected, and phosphorylated by the kinase. B, The histogram depicting the relationship between dephosphorylation and the rephosphorylation efficiency. From 0, 1 h, to 3 h phosphatase treatment, endogenous phosphopeptides identification is decreasing, whereas the newly phosphorylated sites labeled with heavy phosphate are increasing.

In Vitro Kinase Assay with Stable Isotope Labeled ATP—
High background of serine/threonine phosphorylation and potential residual endogenous kinase activities may lead to false positives when it is identified in the kinase reaction sample. Previous KALIP strategies used SILAC method to quantify the change of phosphorylation before and after the kinase reaction. To heighten the sensitivity and specificity of the in vitro kinase reaction, $^{18}$O labeled-ATP (“heavy” ATP) is introduced in the kinase assay (Fig. 2A). In contrast to the naturally existing $[^{16}\text{O}_{4}]$-ATP (“light” ATP), when the heavy ATP is utilized by the kinase for phosphorylation, the substrates will have a phosphate group with three $^{18}$O atoms. Therefore, the newly phosphorylated sites can be qualitatively distinguished from the background in the spectra. The “yes or no” feature of siKALIP improves the sensitivity of in vitro kinase assay and makes the data interpretation more straightforward and re-
Identification of ERK1 Direct Substrates by siKALIP

produced compared with SILAC-based quantitative analysis. Note that P-O bond on the phosphate is quite stable, and it has been demonstrated that $^{18}$O and $^{16}$O exchange did not occur under the catalysis of kinase or harsh conditions during the sample preparation (32).

**Dephosphorylation Significantly Improves the Sensitivity of In Vitro Kinase Assay**—There have been a handful of studies using kinase assay to screen the substrates from whole cell extracts (33, 34). Widespread and high dynamic range of phosphorylation events in whole cell extracts diminishes the sensitivity unless intensive fractionation steps are installed to simplify the sample. We carefully examined the effect of dephosphorylation prior to the kinase reaction. Whole cell extracts prepared from human embryonic kidney 293 (HEK293) cells were treated with phosphatase and monitored as a function of time. These samples were further split for control and kinase assay with ERK1. Following the *in vitro* reaction and assisted by $^{18}$O labeled-ATP, we compared the “heavy” and “light” phosphorylation sites in these samples and a clear correlation between dephosphorylation efficiency and kinase assay sensitivity was observed (Fig. 2B). As expected, a longer dephosphorylation step led to the decreased number of light phosphorylation sites in the control sample. Interestingly, lower background phosphorylation (number of light phosphorylation sites) correlates with higher dephosphorylation level (number of heavy phosphorylation sites) in the kinase assay. Two factors may contribute to the observed correlation: a good percentage of serine/threonine residues were pre-occupied by endogenous phosphorylation so that the stoichiometry of free residues for new phosphorylation events might be too low to be detected in the MS; Another factor might be that large amount of endogenous phosphorylation overwhelms the new phosphorylation events by the kinase (phosphorylated by heavy ATP) during phosphopeptide enrichment and during the LC-MS/MS data acquisition. Taken together, dephosphorylation of cell lysate by a phosphatase increases the sensitivity of *in vitro* kinase assay in the strategy.

**Identification of Direct ERK1 Substrates In Vitro**—We applied the siKALIP strategy to identify novel substrates for ERK1 in HEK293 cells. Protein lysate from the cells was dephosphorylated using an alkaline phosphatase, which was subsequently inactivated by pulse heating. After FSBA treatment, the candidate substrate proteins were then split and incubated with or without purified active ERK1 in the kinase reaction buffer at the presence of $\gamma$-$[^{18}]$O$_4$-ATP. The reaction was ceased by 8 M urea followed by trypsin digestion. Phosphopeptides were enriched and analyzed by mass spectrometry. The existence of heavy phosphate was used as an indicator to determine the direct substrates of ERK1 *in vitro*. Supplemental Fig. S1A showed that the progesterone receptor membrane component 1 (PGRMC1) could not be phosphorylated by ERK1 in our experiment because only the light phosphorylation was detected. In contrast, peptide ASG-QAFELILsPR derived from the known substrate Statmin had the heavy phosphorylation form (Fig. 3A), which was easily distinguished from the background in light phosphorylation. Many novel substrates were identified. For example, BAG family molecular chaperone regulator 3 (BAG3) had two possible sites (S377 and S385) that could be phosphorylated by ERK1 *in vitro* (supplemental Fig. S1B). A previous reported substrate, cortactin, also had two novel phosphorylation sites identified, although neither of these two sites had been dephosphorylated completely (Fig. 3B). In these cases, the coexistence of light and heavy phosphopeptides allowed us to quantify the kinase activity normalized by residual phosphorylation level of substrates.

**Validating ERK1 Specificity for the Substrates Recognition**—From all identified peptides (supplemental Table S1) using two batches of HEK293 cells (400 µg cell lysate for each experiment), a total of 629 unique phosphorylation sites labeled with heavy phosphate group, representing 400 phosphoproteins (supplemental Fig. S2, supplemental Table S2) which were identified exclusively in the kinase reaction samples as *in vitro* ERK1 substrates. The phosphorylation sites favored by protein kinases could be dependent on both primary sequence and the protein structures (35). Compared with protein array-based screening, siKALIP provides a convenient and efficient way to determine kinase specificity by taking advantage of naturally synthesized proteins in an intact cell. We applied the motif analysis (36) to extract the consensus sequence of ERK1 among the peptides with heavy labeled phosphorylation *in vitro* (Fig. 4A). The enrichment of proline residue adjacent to the phosphorylated serine was consistent with reported ERK1 motif according to PhosphoSitePlus (37) (Fig. 4B). Besides the phosphorylation site motif, ERK1 substrates were reported to have a binding sequence named D domain (38) (positively charged residues positioned at three to five residues ahead of a hydrophobic sequence). Using the Scansite (39), we found D domain was also enriched in the *in vitro* substrates compared with the Uniprot Human proteome (Fig. 4C). We further compared our phosphorylation motif and D domain enrichment with other methods previously reported for *in vitro* MAPK substrate screening, including ASKA for ERK2 substrates (40) and high throughput *in vitro* kinase assay for p38 (34). Although three studies shared common features for the consensus motif SP/TP (supplemental Fig. S3A, S3B), the similarity of D domain and ERK1 motif was higher between Carlson’s and ours (Fig. 4C), implying that ERK1 and ERK2 share higher homology compared with p38. This result demonstrated high sensitivity and specificity of our approach in extracting the *in vitro* substrates for serine/threonine kinases.

**Identifying Direct Substrate of ERK1 Under the Physiological Condition by Overlapping In Vitro and In Vivo Phosphoproteins**—Under the *in vitro* reaction condition, kinase specificity is usually compromised because of the high amount of the exogenous kinase and the loss of physiological context. To improve our ability to identify true substrates with a low
false positive rate, we coupled our in vitro result with in vivo ERK1 regulated phosphoproteins from several previous reports (40–44). In these studies, cell permeable inhibitors such as U0126 were applied in the cell culture. Those inhibitors inhibited ERK1 by disrupting the ATP binding to the N-terminal domain of its specific activator MEK1/2 (45). By comparing the control cell versus inhibitor treated cells, the ERK1 dependent phosphoproteins can be retrieved. Based on this approach, Mann’s group identified 98 proteins with 167 phosphorylation sites using SILAC quantitation (41), whereas Ahn’s group identified 35 proteins with 60 phosphorylation sites using label-free quantitation (42). In the same year, Hattori’s group identified 38 proteins using 2D-DIGE, IMAC, and phosphomotif antibodies (43). Recently Thibault and colleagues further expanded the ERK1-dependent repertoire, which contains 155 proteins with 232 phosphorylation sites using a rat cell line (44). In addition, two previously mentioned studies identified in vivo substrates of ERK2 (40) and p38 (34), which also belong to mitogen activation protein kinase (MAPK) family. Overlapping our data with each of those studies was provided in supplemental Fig. S4. Generally speaking, we got 34, 19, 17, and 11 overlapping substrate proteins with Courcelles (44), Pan (41), Old (42), and Kosako’s studies (43), respectively. Considering the high homology between ERK1 and ERK2, we also took 16 ERK2-dependent phosphoproteins identified by Forest group into the analysis and got three common phosphoproteins. Combining those five datasets together, we identified 46 direct phosphorylation sites by ERK1 (Supplemental Table S3).

Validating Novel ERK1 Direct Substrates by In Vitro Kinase Assay—Among the 38 phosphoproteins containing 46 phosphorylation sites overlapping in vitro and in vivo, 14 were known substrates of ERK1 with 18 phosphorylation sites (supplemental Table S3). Note that a novel site T364 was identified on the previously characterized substrate cortactin (46). The remaining 27 phosphorylation sites representing 24 proteins

![Fig. 3. MS spectra of different cases of phosphopeptides identified by LC-MS. A, The known phosphorylation site of ERK1. B, The novel phosphorylation sites of ERK1 identified while the endogenous phosphorylation still exist in high level.](image-url)
were considered to be novel substrates of ERK1. When considering the phosphoprotein regardless of accurate phosphorylation sites, we totally identified 72 proteins that were phosphorylated by ERK1 both in vitro and in vivo (supplemental Table S4).

To gain a global view of ERK1 phosphorylation, we performed gene ontology (GO) to analyze the 72 direct substrates. The cellular components analysis indicated that ERK1 substrates fell into four major categories (Fig. 5A), cytoplasm, nucleus, membrane, and cytoskeleton, a distribution consistent with a previous report (40). In comparison with total human proteome (28,612 proteins) and phosphoproteome (11,479 proteins) having GO annotations from Uniprot and PhosphositePlus, respectively, the proportions of cytoplasm, nucleus, and cytoskeleton were significantly increased (Fig. 5B). Together with the previous in vivo study that spliceosome phosphorylation is ERK1 dependent (44), the enrichment of spliceosomal proteins in our result further suggested that ERK1 probably contributes to the splicing regulation directly through phosphorylation. In addition, the direct ERK1 substrates could be grouped into several important biological processes: gene expression, metabolism, cellular assembly, cell cycle, transport, and signal transduction (Fig. 5C). Among them, gene expression, metabolism, cell assembly, and cell cycle were significantly enriched compared with human proteome (Fig. 5D). ERK1 has been known to regulate gene expression by phosphorylating transcription factors and transporters in both the cytosol and nucleus (47). A handful of substrates involved in gene expression regulation fall in those categories (supplemental Table S5), including some known ERK1 substrates such as nucleoporin (NUP) (43), translocated promoter region (TPR) (48), and heterogeneous nuclear ribonucleoprotein (HNRNP) (49). Besides, many substrates play roles in cytoskeleton organization (supplemental Table S6), such as stathmin (50) and cortactin (51).

The identified substrates are mapped on commonly annotated cellular components and important biological functions (Fig. 6A, supplemental Table S7). Among the newly identified substrate candidates, we selected three for biochemical validation: BAG family molecular chaperone regulator 3 (BAG3), Epidermal growth factor receptor pathway substrate 15 (EPS15), and Cdc42 effector protein 4 (BORG4). BAG3 and EPS15 were immunoprecipitated from HEK293 cell lysates, whereas BORG4 was the recombinant protein purified from bacteria. After $^{32}$P-ATP in vitro kinase assay, all three proteins were confirmed to be the direct substrates of ERK1 by autoradiography (Fig. 6B). EPS15 is involved in cell growth regulation, especially the regulation of mitogenic signals and cell proliferation. It was involved in the internalization of ligand-inducible receptors such as EGFR through a variety of protein-protein interactions (52). PHOSIDA (53) showed that the S662 was phosphorylated during mitosis by a phosphoproteomic study for cell cycle (54). To the best of our knowledge, other than large-scale phosphoproteomic screenings, this is the first report indicating that EPS15 phosphorylation is directly related to ERK1. The significant increase of the phosphorylation and the subsequent biological processes such as protein interaction and membrane translocation suggest that such modification by ERK1 might be essential for EGFR endocytosis. BAG3 executes the anti-apoptotic function by forming a complex with phospholipase C-gamma and Hsp70/
Hsc70 under EGF-regulation (55). Several studies pointed out that S377 and S385 phosphorylation of BAG3 were regulated in the cell cycle (56) and up-regulated under the EGFR stimulation (31). Our research illustrates that direct upstream kinase ERK1 might be critical for BAG3 regulation in both cell cycle and EGFR pathway.

Compared with the BAG3 and EPS15, much less is known about Cdc42 effector protein 4 (BORG4). It may be involved in the cell morphology by inducing the actin filament assembly (57). Although the knowledge of BORG4 function is limited, its S104 phosphorylation appears to be significantly regulated for the cell cycle and EGFR pathway, in agreement with other research groups (31, 56). Elucidation of the direct phosphorylation relationship between ERK1 and BORG4, as well as other identified substrates in the same network, such as CTTNBP2 N-terminal-like protein (CTTNBP2NL), cytoplasmic dynein 1 light intermediate chain 1 (DYNC1LI1), and stathmin (STMN1), advances our knowledge those substrates’ functions and mechanisms in cell organization regulation.

**DISCUSSION**

We present here a proteomic strategy that highlights the use of γ-[^18]O]-ATP in a kinase reaction with a complex protein pool prepared from whole cell extracts. The siKALIP approach offers multiple advantages. First of all, the heavy-ATP has the same biochemical properties as normal ATP and the labeling[^18]O does not exchange with[^16]O; thus, there is a clear relationship between the existence of heavy phosphate and the direct kinase substrate *in vitro*. Second, the dephosphorylation step prior to the kinase reaction efficiently reduces the background phosphorylation and as a result, the sensitivity has been greatly improved while fractionation of samples before mass spectrometric analyses is not needed. Third and finally, the strategy can be applied to any kinase, in...
particular serine/threonine kinase. By overlapping data from *in vitro* kinase assay and global kinase-dependent phosphorylation profiling, the method generates a manageable list of direct kinase substrates with high confidence.

Previous KALIP strategy using regular ATP as the phosphate donor has been successfully applied to identify tyrosine kinase substrates. However, when we extended KALIP to serine/threonine kinase studies, because of widespread serine/threonine phosphorylation, the background phosphorylation in whole cell extracts is extremely high, which has led to low sensitivity in substrates identification. This problem remains even with phosphatase treatment. Taking 100 μg HEK293 cell lysate as an example, 1653 phosphopeptides still existed after 1 h dephosphorylation treatment (Fig. 2B). When we applied KALIP using regular ATP, neither the number of phosphopeptides nor the ion intensity of a majority of phosphopeptides was elevated significantly after kinase reaction across all three experiments (0, 1, and 3 h dephosphorylation treatments), indicating relatively low sensitivity for a global serine/threonine KALIP assay. The problem is even more dominant when the kinase activity is relatively low and the amount of starting material is high. In contrast, using the same 100 μg of HEK293 cell lysates as the starting material, we can easily distinguish 116 newly phosphorylated sites after kinase
Identification of ERK1 Direct Substrates by siKALIP

...assay by database search, without any additional complexity on data analysis such as stochastic ion selection for subtractive MS or statistical significance decision when using ratio change as the criteria. Among the 116 phosphopeptides with heavy phosphate groups, it includes several known substrates of ERK1 such as stathmin and cortactin, as well as a new substrate validated in this study, BAG3. These known and validated substrates were not identified with the original KALIP protocol.

To further demonstrate the superior specificity and sensitivity of siKALIP, we examined two additional kinases, including serine/threonine kinase SnRK2 in Arabidopsis tissues and spleen tyrosine kinase (SYK) in Human DG75 B cells. The results indicated that siKALIP is particularly superior to KALIP for the serine/threonine kinase based on the identification of known and confirmed substrates and higher phosphorylation contrast between samples with or without the kinase. As shown in supplemental Fig. S5A, we observed the advantage of using heavy ATP in the SnRK2 kinase assay. Several known substrates were shown phosphorylation in the control group and thus excluded from the substrate candidate list by the existing KALIP approach. However, the evidence that those proteins could be phosphorylated with SnRK2 in the presence of heavy ATP revealed that they are likely genuine substrates of SnRK2 while they remain at a high level of basal phosphorylation. On the other hand, when comparing existing KALIP and siKALIP, we did not observe a significant improvement on the identification of phosphorylated tyrosine substrates using SYK as the exogenous kinase, in which case both regular and heavy ATP generated around one thousand phosphorylation site in two replicates (supplemental Fig. S5B). We reasoned that it might be due to the fact that endogenous tyrosine phosphorylation is low and the residual tyrosine phosphorylation would not affect MS analysis significantly.

Our list of in vitro ERK1 substrates included 32 biochemically validated ERK1 phosphorylation sites present on 22 substrate proteins, including TPR, STMN, CTTN, c-Jun, LMNA, PXN, EIF4E, and RhoGEF (supplemental Table S8). When high-throughput data was also taken into account, there were 60 phosphorylation sites representing 46 proteins shown to be regulated by ERK1 (supplemental Table S9). Furthermore, to the best of our knowledge, 33 additional proteins can be phosphorylated by ERK1 in vitro (supplemental Table S10). However, the accurate ERK1 phosphorylation sites in those 33 proteins were not clarified since previous studies did not provide phosphorylation sites information or the site localization did not exactly match with our data. Overall, the high recovery of known ERK1 substrates validated the sensitivity of our method.

To further evaluate the sensitivity of isotope labeling in vitro kinase assay during siKALIP procedure, we compared our result with the three most recent studies that performed high throughput in vitro mitogen activated protein kinase (MAPK) substrates screening, including ERK1 kinase assay with phosphospecific antibody (43), analog-sensitive ERK2 kinase assay with SILAC (40), and p38 kinase assay with cell lysate using FSBA and dimethyl labeling quantitation (34). The overlaps of our data with these studies are shown in supplemental Fig. S6. We identified about half of the in vitro substrates from the report of Kosako (43), whereas the overlaps with Carlson (40) and Knight (34) were smaller, which might be attributed to the different organisms chosen and less homology between p38 and ERK1. The results suggest the capability of siKALIP for the in vitro kinase substrates identification and the consistency with other conventional methods.

Although siKALIP is capable of identifying direct substrates of any kinase in high throughput, there are certain conditions we need to be aware of. First, prior to the kinase reaction endogenous kinase activity has to be eliminated using a generic kinase inhibitor such as 5'-(4-fluorosulfonylbenzoyl)adenosine (FSBA). However, complete inhibition of all endogenous kinases is always difficult to achieve so that the remaining kinase activity from the lysate can potentially lead to false positive phosphorylation events. Second, the nature of siKALIP requires sufficient activity of purified kinase in vitro may require a prerequisite activation event (58), cofactor (59), or cellular context (60). Third, a full spectrum of kinase-dependent phosphorylation change in vivo is essential. In this study, Courcelles and coworkers have reported the most comprehensive quantitative phosphoproteomics on ERK1 to our knowledge (44), which led to the most overlapping substrates (34 proteins) with our in vitro kinase reactions. It also highlights the importance of highly specific and efficient kinase inhibition in vivo. Together, only when both in vitro and in vivo datasets are comprehensive and specific can the siKALIP performance on direct substrate identification be maximized.

* This project has been funded in part by an NSF CAREER award CHE-0645020 (W.A.T.), and by a National Institutes of Health grant GM088317 (WAT).

§§ This article contains supplemental Figs. S1 to S7 and Tables S1 to S10.

** To whom correspondence should be addressed: Department of Biochemistry, Purdue University, 175 S University St, West Lafayette, IN 47907. Tel.: 765-494-9605; Fax: 765-494-7897; E-mail: watao@purdue.edu.

ADDITIONAL INFORMATION: Supplementary data set containing 7 figures and 10 tables and annotated sequence spectra supporting identification of phosphorylated peptides can be downloaded are available on the internet through the MCP site. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (61) with the dataset identifier PXD000357.

REFERENCES

Identification of ERK1 Direct Substrates by siKALIP

Identification of ERK1 Direct Substrates by siKALIP

Molecular & Cellular Proteomics 13.11

reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. Nat. Struct. Mol. Biol. 16, 1026–1035


