Proximity Labeling-assisted Identification of Endogenous Kinase Substrates

Tomoya Niinae, Naoyuki Sugiyama, Koshi Imami and Yasushi Ishihama*

Department of Molecular & Cellular BioAnalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, 46–29 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto 606–8501, Japan

Corresponding author: marked as *
Yasushi Ishihama (yishiham@pharm.kyoto-u.ac.jp)

Running title
Identification of kinase substrates

Keywords
Kinase substrate, BioID (proximity-dependent biotin identification), phosphoproteomics

Abbreviations
ACN: Acetonitrile
AP-MS: Affinity-purification mass spectrometry
BioID: Proximity-dependent biotin identification
BirA: Biotin ligase
CK2: Casein kinase 2
GO: Gene Ontology
HAMMOC: Hydroxy acid modified metal oxide chromatography
LC/MS/MS: Liquid chromatography/tandem mass spectrometry
PKA: Protein kinase A
PPI: Protein-protein interaction
PSM: Peptide spectrum match
PWMs: Position weight matrices
TMT: Tandem Mass Tag

Abstract
Mass spectrometry-based phosphoproteomics allows identifying more than 10,000 phosphorylated sites in a single experiment. Despite the fact that enormous phosphosite information has been accumulated in public repositories, protein kinase-

substrate relationships remain largely unknown. Here, we developed a method to identify endogenous substrates of kinase using proximity labeling. We used a proximity-dependent biotin identification approach, called BioID, in combination with kinase-perturbed phosphoproteomics profiling and phosphorylation sequence motifs derived from *in vitro* kinase assay to find molecules that interact with the kinase, are altered in phosphorylation by kinase perturbation, and are directly phosphorylated by the kinase *in vitro*, i.e., the endogenous kinase substrates. This approach was applied to casein kinase 2 and protein kinase A to identify 33 and 52 putative substrates, respectively. In addition, we evaluated the effect of amino acid substitutions around phosphosites of kinase substrates reported in cancer tissues on the kinase-substrate relationship.

Altogether, this study extends our knowledge of kinase-substrate networks by proposing a new approach to identify the endogenous kinase substrates.

Introduction

Protein phosphorylation plays a key role in intracellular signal transduction and regulates various biological processes, including cell proliferation and differentiation. Mass spectrometry (MS)-based phosphoproteomics has allowed us to identify thousands phosphorylated sites in single experiments (1–3). Despite the fact that enormous phosphosite information has been accumulated in public repositories (4–7), protein kinase-substrate relationships remain largely unknown both *in vitro* and *in vivo* (8).

In vitro kinase assay is one of the most classical and widely used approaches to identify kinase substrates (9–13). We recently reported a total of 175,574 substrates for 385 kinases using *in vitro* kinase reaction with protein extract from human cells, followed by phosphopeptide enrichment and liquid chromatography-tandem mass spectrometry (LC/MS/MS) analyses (14, 15). While this method successfully identified *in vitro* substrates and uncovered a variety of consensus motifs for kinase substrates, we still lack information on endogenous kinase substrates as physiological conditions within cells are not considered in these studies. Perturbation of kinase activity in living cells through drug treatment (16, 17) or knocking down/out a specific kinase (18, 19) allows us to monitor consequent changes in the phosphorylation level *in vivo*. However, these approaches would also indirectly affect downstream kinases, which would interfere with the identification of direct substrates (20). To overcome these issues, we and others employed *in vitro* substrate information or the sequence motif in addition to the kinase-perturbed phosphoproteomics using living cells. Consequently, the

endogenous substrate candidates of the target kinases were successfully obtained for protein kinase A (PKA), spleen tyrosine kinase (Syk) and Abelson tyrosine kinase (ABL) (21–23).

Although the combined approach of *in vitro* substrate information and kinase-perturbed phosphoproteomic profiling has identified putative substrates in some cases, this strategy cannot distinguish between substrates of downstream kinases that phosphorylate motifs similar to the target kinase. Therefore, we reasoned that an additional layer of information on kinase-protein interactions including transient interactors should allow us to identify more confident endogenous substrates. Indeed, computational analyses have demonstrated that using protein-protein interaction-derived networks and kinase-specific phosphorylation sequence motifs improved prediction of the substrate specificity (21–23). However, these studies relied on public protein interaction databases in which protein interactions were measured using different cell types in different biological contexts. Furthermore, protein interactions were determined through classical approaches such as affinity-purification mass spectrometry (AP-MS) and yeast two-hybrid, by which transient interactors are generally not captured. To determine the endogenous substrates of target kinase, interactome analysis which capture not only stable interactions but also transient interactions, should be performed with kinase-perturbed phosphoproteomic profiling in parallel using the same sample.

Recently, proximity labeling approaches, such as proximity-dependent biotin identification (BioID) (24) and an engineered ascorbate peroxidase (APEX) (25), have been developed to capture transient interactions including kinase-substrate interactions (26–28). BioID is based on biotinylation of proteins proximal (~10 nm) to a mutant biotin ligase (BirA*)-fused protein of interest, and the biotinylated proteins can be captured and identified with streptavidin pulldown and LC/MS/MS. Thus, BioID is an ideal tool for interactome analysis to globally capture transient interactions.

In this study, we combined BioID-based interactome analysis with kinase-perturbed phosphoproteomic profiling and substrate motif analysis to establish a common framework for the systematic analysis of the endogenous kinase-substrate relationship.

EXPERIMENTAL PROCEDURES

Cell culture

HEK293T cells were provided by the RIKEN BRC through the National BioResource Project of the MEXT, Japan. HEK293T cells and HeLa cells (HSRRB, Osaka, Japan) were cultured in DMEM (Fujifilm Wako, Osaka, Japan) containing 10%

fetal bovine serum (Thermo Fisher Scientific, Waltham, MA) and 100 µg/mL penicillin/streptomycin (Fujifilm Wako).

Cloning of BirA*-kinase expression vectors

pDEST-pcDNA5-BirA*-FLAG N-term was a gift from Dr. Anne-Claude Gingras (Lunenfeld-Tanenbaum Research Institute, Canada). The entry clones pENTR221 CK2A1 and pENTR221 PRKACA were purchased from DNAFORM and the RIKEN BRC through the National BioResource Project of the MEXT/AMED, Japan, respectively. CK2A1 or PRKACA coding sequences were cloned into the destination vector pDEST-pcDNA5-BirA*-FLAG N-term with LR clonase 2 (Thermo Fisher Scientific) via Gateway system.

BioID

For kinase interactome experiments, HEK293T cells in a 10 cm dish were transfected with 40 µL 1.0 mg/mL polyethylenimine (Polysciences, Warrington, PA) and 15 µg plasmid and incubated for 24 h. For negative control, HEK293T cells in 10 cm dish were transfected with 40 µL 1.0 mg/mL polyethylenimine for 24 h. all experiments were performed in triplicates. Then, cells were incubated for 24 h in culture medium containing 50 µM biotin (Fujifilm Wako). Cells were washed and harvested with ice-cold PBS.

Drug treatment

HEK293T cells were treated with DMSO as control, 10 µM CX-4945 (ApexBio, Houston, TX) or 50 µM forskolin (Fujifilm Wako) for 1 hr. Biological triplicates were performed.

Sample preparation for biotinylated protein identification

Cells were washed and harvested with ice-cold PBS. The proteins were extracted by the RIPA buffer (50 mM Tris-HCl (pH 7.2), 150 mM, NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0.1% SDS, protease inhibitors cocktail (Sigma-Aldrich, St. Louis, MO), and 1% sodium deoxycholate). Proteins were rotated with 300 µg streptavidin magnetic beads (Thermo Fisher Scientific) for 3 hr at 4 °C. After incubation, beads were washed with RIPA buffer three times and 50 mM ammonium bicarbonate buffer 3 times and suspended in 200 µL of 50 mM ammonium bicarbonate buffer. Then, captured proteins were reduced by 10 mM DTT for 30 min, alkylated with 50 mM iodoacetamide for 30 min in the dark, and digested with Lys-C (w/w 1:100) for 3 h, followed by trypsin digestion (w/w 1:100) overnight at 37 °C, on beads. Then, the peptides were desalted using StageTip (29) with SDB-XC Empore disk membranes

(GLSciences, Tokyo, Japan) and suspended in the loading buffer (0.5% TFA and 4% acetonitrile (ACN) for subsequent LC/MS/MS analyses.

Sample preparation for phosphoproteome analysis

Cells were washed and harvested with ice-cold PBS. The proteins were extracted by the phase transfer surfactant (30) using a lysis buffer (12 mM sodium deoxycholate (Fujifilm Wako), 12 mM sodium N-lauroylsarcosinate (Fujifilm Wako), 100 mM Tris-HCl (pH 9.0), containing protein phosphatase inhibitor cocktail 1 and 2 (Sigma-Aldrich) and protease inhibitors (Sigma-Aldrich). Protein amount was confirmed with a BCA protein assay kit and reduced with 10 mM DTT for 30 min, alkylated with 50 mM iodoacetamide for 30 min in the dark. After reduction and alkylation, proteins were digested with Lys-C (w/w 1:100) for 3 h, followed by trypsin digestion (w/w 1:100) overnight at 37 °C. Then, the peptides were desalted using SDB-XC StageTip.

Phosphopeptides were enriched from 100 µg of tryptic digested peptides with TiO_2-based hydroxy acid modified metal oxide chromatography (HAMMOC) (31) and eluted by 0.5% piperidine. Phosphopeptides were labeled with Tandem Mass Tag (TMT) (Thermo Fisher Scientific), desalted using SDB-XC StageTips and suspended in the loading buffer (0.5% TFA and 4% ACN) for subsequent LC/MS/MS analyses.

***In vitro* kinase assay using cell extracts**

Cells were washed and harvested with ice-cold PBS. Proteins were extracted from HeLa cells by the phase transfer surfactant. The buffer was replaced with 40 mM Tris-HCl (pH7.5) by ultrafiltration using an Amicon Ultra 10K at 14,000 g and 4 °C. Protein amount was confirmed with a BCA protein assay kit and distributed to each of 100 µg aliquot. As described previously (14), proteins were dephosphorylated by TSAP, reacted with recombinant kinase and digested by Lys-C (w/w 1:100) and trypsin (w/w 1:100) as described above. Then, phosphopeptides were enriched from tryptic digested peptides with HAMMOC and desalted using SDB-XC StageTips and suspended in the loading buffer (0.5% TFA and 4% ACN) for subsequent LC/MS/MS analyses.

***In vitro* kinase assay using synthetic peptides**

A mixture of synthetic peptides (SynPeptide, Shanghai, China) (10 pmol each) was reacted with 0.5 µg of each recombinant kinase in 100 µL of a kinase reaction buffer (40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 1 mM ATP) at 37°C for 3 h. The reaction was quenched by adding 10 µL 10% TFA. Then, the peptides were desalted using SDB-XC StageTip and suspended in the loading buffer (0.5% TFA and 4% ACN) for subsequent LC/MS/MS analyses.

NanoLC-MS/MS analyses

NanoLC/MS/MS analyses were performed on an Orbitrap Fusion Lumos (Thermo Fisher Scientific) or a Q Exactive (Thermo Fisher Scientific) only for synthetic peptides samples, which is connected to an Ultimate 3000 pump (Thermo Fisher Scientific) and an HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland). Peptides were separated by a self-pulled needle column (150 mm length x 100 µm ID, 6 µm opening) packed with Reprosil-C18 AQ 3 µm reversed phase material (Dr. Maisch, Ammerbuch, Germany). The flow rate was set to 500 nL/min. The mobile phase consisted of (A) 0.5% acetic acid and (B) 0.5% acetic acid in 80% acetonitrile. Three-step linear gradients of 5-10% B in 5 min, 10-40% B in 60 min (for short gradient) or 100 min (for long gradient), 40-100% B in 5 min were employed.

For BioID samples and phosphopeptides obtained with *in vitro* kinase reaction, the MS scan range was *m/z* 300-1500. MS scans were performed by the Orbitrap with *r* =120,000 and subsequent MS/MS scans were performed by the Orbitrap with *r* = 1,5000. Auto gain control for MS was set to 4.00 X 10^5 and that for MS/MS was set to 5.00 X 10^4. The HCD was set to 30.

For TMT labeled samples, synchronous precursor selection-MS3 (SPS-MS3) (32) was performed. The MS scan range was *m/z* 375-1500. MS scans were performed by the Orbitrap with *r* =120,000, MS/MS scans were performed by the Ion Trap with Turbo mode and MS3 scans were performed by the Orbitrap with *r* = 1,5000. Auto gain control for MS was set to 4.00 X 10^5, that for MS/MS was set to 1.00 X 10^4 and that for MS3 was set to 5.00 X 10^4. The CID was set to 35.

For synthetic peptides samples, the MS scan range was *m/z* 300-1500. MS scans were performed by the Orbitrap with *r* =70,000 and subsequent MS/MS scans were performed by the Orbitrap with *r* = 17,500. Auto gain control for MS was set to 3.00 X 10^6 and that for MS/MS was set to 1.00 X 10^5. The HCD was set to 27.

Database searching

For BioID experiments, the raw MS data files were analyzed by MaxQuant v1.6.2.3 (33). Peptides and proteins were identified by means of automated database searching using Andromeda against human SwissProt Database (version 2017-04, 20,199 protein entries) with a precursor mass tolerance of 20 ppm for first search and 4.5 ppm for main search and a fragment ion mass tolerance of 20 ppm. Cysteine carbamidomethylation was set as a fixed modification. Methionine oxidation and acetyl on protein N-term were set as variable modification. The search results were filtered with FDR < 1% at peptide spectrum match (PSM) and protein levels.

For kinase perturbation experiments, the raw MS data files were analyzed by ProteoWizard (34), ProteomeDiscoverer v1.3.0.339 and MaxQuant v1.6.2.10 to create peak lists based on the recorded fragmentation spectra. Peptides and proteins were

identified by means of automated database searching using Mascot v2.4 (Matrix Science, London) against human SwissProt Database (version 2017-04, 20,199 protein entries) with a precursor mass tolerance of 5 ppm, a fragment ion mass tolerance of 20 ppm. Cysteine carbamidomethylation and TMT tags on lysine and peptide N-termini were set as fixed modifications. Methionine oxidation and phosphorylation on serine, threonine and tyrosine were set as variable modifications. The search results were filtered with FDR < 1% at peptide level.

For *in vitro* kinase reaction using HeLa cells, the raw MS data files were analyzed by ProteoWizard to create peak lists based on the recorded fragmentation spectra. Peptides and proteins were identified by means of automated database searching using Mascot v2.4 (Matrix Science, London) against human SwissProt Database (version 2017-04, 20,199 protein entries) with a precursor mass tolerance of 5 ppm, a fragment ion mass tolerance of 20 ppm. Cysteine carbamidomethylation was set as a fixed modification. Methionine oxidation and phosphorylation on serine, threonine and tyrosine was set as variable modification. The search results were filtered with FDR < 1% at peptide level. As described previously (35), we also used additional criterion that at least three successive y- or b-ions with a further two or more y-, b- and/or precursor-origin neutral loss ions were observed, based on the error-tolerant peptide sequence tag concept (36). Phosphosite localization was confirmed with a site-determining ion combination method (37). This method is based on the presence of site-determining y- or b-ions in the peak lists of the fragment ions, which unambiguously identify the phosphosites. Only singly phosphorylated peptides with confident phosphosites were accepted for the following analysis.

Data analysis

For BioID samples, the peak area of each peptide in MS1 was quantified using MaxQuant. Missing values were imputed with values representing a normal distribution around the detection limit of the mass spectrometer. A two-tailed Welch's t-test was performed comparing the BirA*-kinase group to the control group.

For TMT samples, the peak intensities of reporter ions were obtained from MS3 scan. Missing values were imputed with values representing a normal distribution around the detection limit of the mass spectrometer. The ratio of drug-treated to control were logged (base 2) for each phosphopeptide, and averaged to each phosphosite.

For synthetic peptides, phosphorylation ratio was calculated as follows:

$$Phosphorylation\ ratio\ = 1 - \frac{peak\ area_{Nonphosphopeptide\ (w/\ kinase)}}{peak\ area_{Nonphosphopeptide\ (w/o\ kinase)}}$$

The protein-protein interaction (PPI) network analysis was performed with STRING version 11.0 (38) and the medium-confidence PPI network (score = 0.4) was used. GO analysis was performed with the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8(39, 40). Background was set as Homo sapiens.

Motif score

The probability of observing residue x in position i is computed as follows:

$$p(x,i) = \frac{f_{x,i} + c(x)}{N},$$

$$c(x) = p(x) \cdot \epsilon$$

where $f_{x,i}$ is the frequency of observing residue x at position i and N is the total number of sequences. $c(x)$ is a pseudo count function which is computed as the probability of observing residue b in the proteome, $p(x)$, multiplied by ε, defined as the square root of the total number of sequences used to train the PWM. This avoids infinite values when computing logarithms. Probabilities are then converted to weights as follows:

$$w_{x,i} = log_2 \frac{p(x,i)}{p(x)}$$

where $p(x)$ = background probability of amino acid x; $p(x,i)$ = corrected probability of amino acid x in position i; $W_{x,i}$ = PWM value of amino acid x in position i. Given a sequence q of length l, a score λ is then computed by summing $\log_2$ weights:

$$\lambda = \sum_{i=1}^{l} W_{q_i,i}$$

where q_i is the ith residue of q. In this study, the score was computed using the flanking ±7 residues surrounding each phosphosite.

Results

Defining the kinase interacting proteins using BioID

To identify kinase interacting proteins, we first performed BioID experiments (Fig. 1A). We selected CK2 and PKA as their functions, localizations and putative substrates have been relatively well investigated (8) . BirA*-fused kinases ($BirA_k$) were individually transfected into HEK293T cells, and non-transfected cells were used as negative control. Of note, we observed broad-spectrum of biotinylated proteins by streptavidin blotting for the $BirA_k$ expressing cells, while no detectable signals were observed for the control cells (Fig. S1A), demonstrating the validity of our BioID

experiments. After the transfection and biotinylation, biotinylated proteins were enriched with streptavidin beads, digested into tryptic peptides, analyzed by nanoLC/MS/MS and quantified using label-free quantification (41). As a result, 2,218 proteins and 1,795 proteins were quantified in at least two of the three replicates in the CK2- and PKA-expressing cells, respectively. Among them, 574 and 518 proteins were considered as the interacting proteins of CK2 and PKA, respectively, after applying the cut-off values (Fig. 1B, Table S1). Importantly, CSNK2B and PRKAR1A, which form the heteromeric complexes with the corresponding kinase, were significantly enriched as the kinase interactors (Fig. 1B), suggesting that the tagged kinases form the heteromeric complexes *in vivo* and are likely to function as the endogenous kinases.

To benchmark our method, we evaluated the biotinylation of previously known interactors for CK2 and PKA (38). We found that the known interactors exhibited significantly higher ratios of $BirA_{R}$/control compared to all proteins quantified in this study (Fig. S1B). Then, we mapped CK2- or PKA-interacting proteins (n=574 and n=518, respectively) on the STRING (38) -based protein interaction networks. We found that 395 proteins (correspond to 69%, p-value < 10^{-16}) and 371 proteins (72%, p-value < 10^{-16}) composed the single large networks (that is, any given pairs of two proteins are interconnected directly and/or indirectly.) (Fig. 1C). This indicates that our BioID experiments captured biologically meaningful and proximal proteins to the target kinases, and that the remaining approximately 30% of the interactors are likely to be transient interactors that were difficult to detect by other methods. Finally, we performed Gene Ontology (GO) enrichment analyses for the sets of the kinase-interacting proteins (Fig. S1C). We found that GO terms related to "cellular proliferation" and "apoptosis" were enriched in the CK2 interacting proteins, consistent with the previously known functionality and biology of CK2 family (42). For the PKA interacting proteins, we observed the term "cell-cell adhesion", which is also in line with PKA's function (43).

Altogether, our data are indeed of high confidence and provide a rich resource of CK2 and PKA interactomes including both stable and transient interactors.

Phosphoproteomic profiling responses to perturbations of kinases

Having established the CK2 and PKA interactomes, we next sought to perform kinase-perturbed phosphoproteome profiling to identify phosphosites regulated by the target kinases (Fig. 2A). To this end, HEK293T cells were treated with either an ATP-competitive CK2 inhibitor, CX-4945, also known as Silmitasertib (44) or a PKA activator, forskolin (45). Following protein extraction and tryptic digestion, phosphopeptides were enriched with titania chromatography (31), labeled with TMT

reagents and analyzed by nanoLC/MS/MS. To simplify the interpretation of kinase-phosphosite relationship, we only considered mono-phosphopeptides for subsequent analyses; as a result, 9,313 and 9,140 mono-phosphosites were quantified in triplicate from CX-4945- and forskolin-treated cells, respectively (Table S2). To ensure our results, down-regulation of known CK2 substrates such as TOP2A (S1377), LIG1 (S66) and MRE11 (S649) as well as up-regulation of known PKA substrates such as CAD (S1406), NDE1 (S306) and FLNA (S2152) were confirmed (Fig. S2A). Furthermore, expected sequence motifs for acidophilic CK2 and basophilic PKA were clearly enriched in perturbed phosphosites (2-fold change by CK2 inhibitor or PKA activator with p-value < 0.05) (Fig. S2B), demonstrating that high quality and meaningful data were obtained in this experiment.

To combine two results obtained from the phosphoproteomic and BioID experiments, we mapped the identified phosphorylation sites onto the kinase interacting proteins; 1,292 and 1,322 phosphosites were attributed to the 305 and 313 interacting proteins with CK2 and PKA, respectively (Fig. 2B). We then set the cut-off, 2-fold change on average (n=3) with p-value < 0.05, to further select the regulated phosphosites, resulting in 67 and 74 phosphosites for CK2 and PKA substrates, respectively.

Sequence motif analysis

Our kinase-perturbed phosphoproteomic analysis identified the regulated sites, which could be due to indirect effects such as perturbation of downstream kinases of the target kinases or off-target effects of the drugs for perturbation. Thus, the next step is to select phosphosites that match the consensus phosphorylation motif of the kinase, which has been shown to be a useful predictor of kinase identification (35).

Position weight matrices (PWMs) (46) represent the signatures of the sequences flanking the phosphosites targeted by a given kinase (Fig. S3). Here columns in the matrix represent relative positions from phosphosites, while rows represent residues. The values in the matrix are log2-transformed probabilities of a residue's occurrence at a position. These probabilities are used to predict putative substrates (see methods). This approach has been implemented in several kinase-substrate prediction methods and applied to *in vivo* kinase−substrate discovery. We built the PWM based on thousands of substrates identified in our previous study (47) and this study (Fig. 2C), and calculated the motif scores for the substrate candidates. Consequently, 33 CK2 substrates and 52 PKA substrates met the criteria (motif score > 2, Table 1). These substrates include the well-known substrates such as TOP2A (S1377) for CK2 and FLNA (S2152) for PKA. Note that only two proteins (NAP1L1, TOP2A) are annotated as CK2 interactors in STRING, while none of them are described as PKA interactors

in STRING, meaning that transient interactors captured by BioID might be survived through these criteria. The GO enrichment analysis revealed that RNA splicing-related proteins were enriched as CK2 substrates, which is consistent with CK2 being an RNA splicing-related kinase (48) , while cell adhesion-related molecules have been identified as PKA substrates since PKA is involved in the cell-cell adhesion pathway (46).

Missense mutations near phosphosites of CK2 and PKA substrates

Having discovered putative direct substrates of CK2 and PKA, we next sought to assess the relationship between the kinase and substrates in the context of cancer biology (49, 50). The effect of amino acid substitutions in substrates on the kinase-substrate relationship has been analyzed using kinase substrate sequence specificity (51, 52). Information on amino acid substitutions that occur within seven residues around the phosphosites in Table 1 was extracted from cancer genomics database cBioPortal (53, 54). As a result, 121 missense mutations from 33 CK2 substrates and 189 missense mutations from 52 PKA substrates were found. For substrate sequences containing these mutations, we used PWM scores to predict phosphorylation preference by kinases and compared them to wild-type sequences (Fig. S4). Then, wild-type/mutant substrate pairs (five substrate pairs for CK2 and one pair for PKA) that showed potential loss of phosphorylation by mutation were selected for *in vitro* kinase assays using synthetic peptides (Fig. 4). As a result, the mutation-induced reduction in phosphorylation stoichiometry was observed for 5 out of the 6 pairs. These results indicate that missense mutations near phosphosites of substrates negatively affect phosphorylation by CK2 or PKA, thereby affect downstream signaling networks in corresponding cancer cells.

Discussion

Kinase-substrate relationships are one of the key information to understand intracellular signaling networks. To our knowledge, this study is the first challenge to identify endogenous kinase substrates by combining three approaches such as proximity labeling, kinase-perturbed phosphoproteomics, and motif preference analysis. Using this approach, we were able to identify novel endogenous substrates as well as known endogenous substrates. This approach has multiple advantages. First, BioID is able to identify transient interactors that could not be covered by conventional methods. Second, motif preference analysis can be applied to any sequence, as it scores matches against substrate motifs rather than against substrates obtained in *in vitro* kinase assay. Finally, this combined approach (BioID, kinase-

perturbed phosphoproteomics, and motif analysis) can be performed for any kinase as a universal method.

One of the applications of this approach is the identification of the endogenous substrate of the kinase linking to the identification of kinase substrate mutations in cancer. We demonstrated the potential impact of the oncogenic mutations and the usefulness of motif analysis for screening. These results provide new possibilities for large-scale discovery of endogenous kinase substrates.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral. proteomexchange.org) via the jPOST partner repository (http://jpostdb.org) (55) with the data set identifier PXD019664.

ACKNOWLEDGMENTS

We would like to thank Dr. Anne-Claude Gingras (Lunenfeld-Tanenbaum Research Institute) for providing BirA* constructs and members of Department of Molecular & Cellular BioAnalysis for fruitful discussion. This work was supported by JST Strategic Basic Research Program CREST (No. 18070870), AMED Advanced Research and Development Programs for Medical Innovation CREST (18068699) and JSPS Grants-in-Aid for Scientific Research No. 17H03605 to Y.I., No. 20H03241 to K.I. and 20H04845 to N.S.

Author contributions

T.N, K.I., N.S. and Y.I. designed research. T.N. performed research and analyzed data. T.N, K.I. and Y.I. wrote the paper.

1. Humphrey, S. J., Babak Azimifar, S., and Mann, M. (2015) High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nature Biotechnology* 33, 990–995

2. Bekker-Jensen, D. B., Bernhardt, O. M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C. D., Reiter, L., and Olsen, J. V. (2020) Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* 11, 787

3. Sugiyama, N. (2020) Mass Spectrometry-Based Discovery of *in vitro* Kinome Substrates. *Mass Spectrometry* 9, A0082–A0082

4. Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512–20

5. Ochoa, D., Jarnuczak, A. F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., Kleefeldt, A. A., Hill, A., Garcia-Alonso, L., Stein, F., Krogan, N. J., Savitski, M. M., Swaney, D. L., Vizcaíno, J. A., Noh, K.-M., and Beltrao, P. (2020) The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* 38, 365–373

6. Dinkel, H., Chica, C., Via, A., Gould, C. M., Jensen, L. J., Gibson, T. J., and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res.* 39, D261–7

7. Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D. S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C. J., Kanth, S., Ahmed, M., Kashyap, M. K., Mohmood, R., Ramachandra, Y. L., Krishna, V., Rahiman, B. A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., and Pandey, A. (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res.* 37, D767–72

8. Needham, E. J., Parker, B. L., Burykin, T., James, D. E., and Humphrey, S. J. (2019) Illuminating the dark phosphoproteome. *Sci. Signal.* 12,

9. Knebel, A. (2001) A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *The EMBO Journal* 20, 4360–4369

10. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R. R., Schmidt, M. C., Rachidi, N., Lee, S.-J., Mah, A. S., Meng, L., Stark, M. J. R., Stern, D. F., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P. F., and Snyder, M. (2005) Global analysis of protein phosphorylation in yeast. *Nature* 438, 679–684

11. Wang, C., Ye, M., Bian, Y., Liu, F., Cheng, K., Dong, M., Dong, J., and Zou, H. (2013) Determination of CK2 specificity and substrates by proteome-derived peptide libraries. *J. Proteome Res.* 12, 3813–3821

12. Newman, R. H., Hu, J., Rho, H.-S., Xie, Z., Woodard, C., Neiswinger, J., Cooper, C., Shirley, M., Clark, H. M., Hu, S., Hwang, W., Jeong, J. S., Wu, G., Lin, J., Gao, X., Ni, Q., Goel, R., Xia, S., Ji, H., Dalby, K. N., Birnbaum, M. J., Cole, P. A., Knapp, S., Ryazanov, A. G., Zack, D. J., Blackshaw, S., Pawson, T., Gingras, A.-C., Desiderio, S., Pandey, A., Turk, B. E., Zhang, J., Zhu, H., and Qian, J. (2013) Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol.* 9, 655

13. Sugiyama, N., and Ishihama, Y. (2016) Large-scale profiling of protein kinases for cellular signaling studies by mass spectrometry and other techniques. *J. Pharm. Biomed. Anal.* 130, 264–272

14. Imamura, H., Sugiyama, N., Wakabayashi, M., and Ishihama, Y. (2014) Large-Scale Identification of Phosphorylation Sites for Profiling Protein Kinase Selectivity. *Journal of Proteome Research* 13, 3410–3419

15. Sugiyama, N., Imamura, H., and Ishihama, Y. (2019) Large-scale Discovery of Substrates of the Human Kinome. *Sci. Rep.* 9, 10503

16. Imami, K., Sugiyama, N., Imamura, H., Wakabayashi, M., Tomita, M., Taniguchi, M., Ueno, T., Toi, M., and Ishihama, Y. (2012) Temporal profiling of lapatinib-suppressed phosphorylation signals in EGFR/HER2 pathways. *Mol. Cell. Proteomics* 11, 1741–1757

17. Pan, C., Olsen, J. V., Daub, H., and Mann, M. (2009) Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol. Cell. Proteomics* 8, 2796–2808

18. Kim, J.-Y., Welsh, E. A., Oguz, U., Fang, B., Bai, Y., Kinose, F., Bronk, C., Remsing Rix, L. L., Beg, A. A., Rix, U., Eschrich, S. A., Koomen, J. M., and Haura, E. B. (2013) Dissection of TBK1 signaling via phosphoproteomics in lung cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12414–12419

19. Chen, Z., Lei, C., Wang, C., Li, N., Srivastava, M., Tang, M., Zhang, H., Choi, J. M., Jung, S. Y., Qin, J., and Chen, J. (2019) Global phosphoproteomic analysis reveals ARMC10 as an AMPK substrate that regulates mitochondrial dynamics. *Nat. Commun.* 10, 104

20. Bodenmiller, B., Wanka, S., Kraft, C., Urban, J., Campbell, D., Pedrioli, P. G., Gerrits, B., Picotti, P., Lam, H., Vitek, O., Brusniak, M.-Y., Roschitzki, B., Zhang, C., Shokat, K. M., Schlapbach, R., Colman-Lerner, A., Nolan, G. P., Nesvizhskii, A. I.,

Peter, M., Loewith, R., von Mering, C., and Aebersold, R. (2010) Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci. Signal.* 3, rs4

21. Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A. T. M., Jørgensen, C., Miron, I. M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J. G., Samson, L. D., Woodgett, J. R., Russell, R. B., Bork, P., Yaffe, M. B., and Pawson, T. (2007) Systematic Discovery of In Vivo Phosphorylation Networks. *Cell* 129, 1415–1426

22. Yachie, N., Saito, R., Sugiyama, N., Tomita, M., and Ishihama, Y. (2011) Integrative features of the yeast phosphoproteome and protein-protein interaction map. *PLoS Comput. Biol.* 7, e1001064

23. Wagih, O., Sugiyama, N., Ishihama, Y., and Beltrao, P. (2016) Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks. *Mol. Cell. Proteomics* 15, 236–245

24. Roux, K. J., Kim, D. I., Raida, M., and Burke, B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* 196, 801–810

25. Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., and Ting, A. Y. (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* 339, 1328–1331

26. Cutler, J. A., Tahir, R., Sreenivasamurthy, S. K., Mitchell, C., Renuse, S., Nirujogi, R. S., Patil, A. H., Heydarian, M., Wong, X., Wu, X., Huang, T.-C., Kim, M.-S., Reddy, K. L., and Pandey, A. (2017) Differential signaling through p190 and p210 BCR-ABL fusion proteins revealed by interactome and phosphoproteome analysis. *Leukemia* 31, 1513–1524

27. Liu, X., Salokas, K., Tamene, F., Jiu, Y., Weldatsadik, R. G., Öhman, T., and Varjosalo, M. (2018) An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* 9, 1188

28. Méant, A., Gao, B., Lavoie, G., Nourreddine, S., Jung, F., Aubert, L., Tcherkezian, J., Gingras, A.-C., and Roux, P. P. (2020) Proteomic Analysis Reveals a Role for RSK in p120-catenin Phosphorylation and Melanoma Cell-Cell Adhesion. *Mol. Cell. Proteomics* 19, 50–64

29. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906

30. Masuda, T., Sugiyama, N., Tomita, M., and Ishihama, Y. (2011) Microscale Phosphoproteome Analysis of 10 000 Cells from Human Cancer Cell Lines. *Analytical Chemistry* 83, 7698–7703

31. Sugiyama, N., Masuda, T., Shinoda, K., Nakamura, A., Tomita, M., and Ishihama, Y. (2007) Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol. Cell. Proteomics* 6, 1103–1109

32. McAlister, G. C., Nusinow, D. P., Jedrychowski, M. P., Wühr, M., Huttlin, E. L., Erickson, B. K., Rad, R., Haas, W., and Gygi, S. P. (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* 86, 7150–7158

33. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 26, 1367–1372

34. Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S. L., Nuwaysir, L. M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E. W., Moritz, R. L., Katz, J. E., Agus, D. B., MacCoss, M., Tabb, D. L., and Mallick, P. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 30, 918–920

35. Imamura, H., Wagih, O., Niinae, T., Sugiyama, N., Beltrao, P., and Ishihama, Y. (2017) Identifications of Putative PKA Substrates with Quantitative Phosphoproteomics and Primary-Sequence-Based Scoring. *Journal of Proteome Research* 16, 1825–1830

36. Mann, M., and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66, 4390–4399

37. Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* 153, 1161–1174

38. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. von (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613

39. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57

40. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13

41. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526

42. Litchfield, D. W. (2003) Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem. J* 369, 1–15

43. Whittard, J. D., and Akiyama, S. K. (2001) Positive regulation of cell-cell and cell-substrate adhesion by protein kinase A. *J. Cell Sci.* 114, 3265–3272

44. Siddiqui-Jain, A., Drygin, D., Streiner, N., Chua, P., Pierre, F., O'Brien, S. E., Bliesath, J., Omori, M., Huser, N., Ho, C., Proffitt, C., Schwaebe, M. K., Ryckman, D. M., Rice, W. G., and Anderes, K. (2010) CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signaling and exhibits antitumor efficacy. *Cancer Res.* 70, 10288–10298

45. Seamon, K. B., Daly, J. W., Metzger, H., de Souza, N. J., and Reden, J. (1983) Structure-activity relationships for activation of adenylate cyclase by the diterpene forskolin and its derivatives. *J. Med. Chem.* 26, 436–439

46. Wasserman, W. W., and Sandelin, A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* 5, 276–287

47. Imamura, H., Sugiyama, N., Wakabayashi, M., and Ishihama, Y. (2014) Large-Scale Identification of Phosphorylation Sites for Profiling Protein Kinase Selectivity. *Journal of Proteome Research* 13, 3410–3419

48. Trembley, J. H., Tatsumi, S., Sakashita, E., Loyer, P., Slaughter, C. A., Suzuki, H., Endo, H., Kidd, V. J., and Mayeda, A. (2005) Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. *Mol. Cell. Biol.* 25, 1446–1457

49. Caretta, A., and Mucignat-Caretta, C. (2011) Protein kinase a in cancer. *Cancers* 3, 913–926

50. Chua, M. M. J., Ortega, C. E., Sheikh, A., Lee, M., Abdul-Rassoul, H., Hartshorn, K. L., and Dominguez, I. (2017) CK2 in Cancer: Cellular and Biochemical Mechanisms and Potential Therapeutic Target. *Pharmaceuticals* 10,

51. Wagih, O., Reimand, J., and Bader, G. D. (2015) MIMP: predicting the impact of mutations on kinase-substrate phosphorylation. *Nat. Methods* 12, 531–533

52. Reimand, J., Wagih, O., and Bader, G. D. (2013) The mutational landscape of phosphorylation signaling in cancer. *Sci. Rep.* 3, 2651

53. Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013) Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling* 6, l1–pl1

54. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012) The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* 2, 401–404

55. Moriya, Y., Kawano, S., Okuda, S., Watanabe, Y., Matsumoto, M., Takami, T., Kobayashi, D., Yamanouchi, Y., Araki, N., Yoshizawa, A. C., Tabata, T., Iwasaki, M., Sugiyama, N., Tanaka, S., Goto, S., and Ishihama, Y. (2019) The jPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Res.* 47, D1218–D1224

56. Keilhauer, E. C., Hein, M. Y., and Mann, M. (2015) Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol. Cell. Proteomics* 14, 120–135

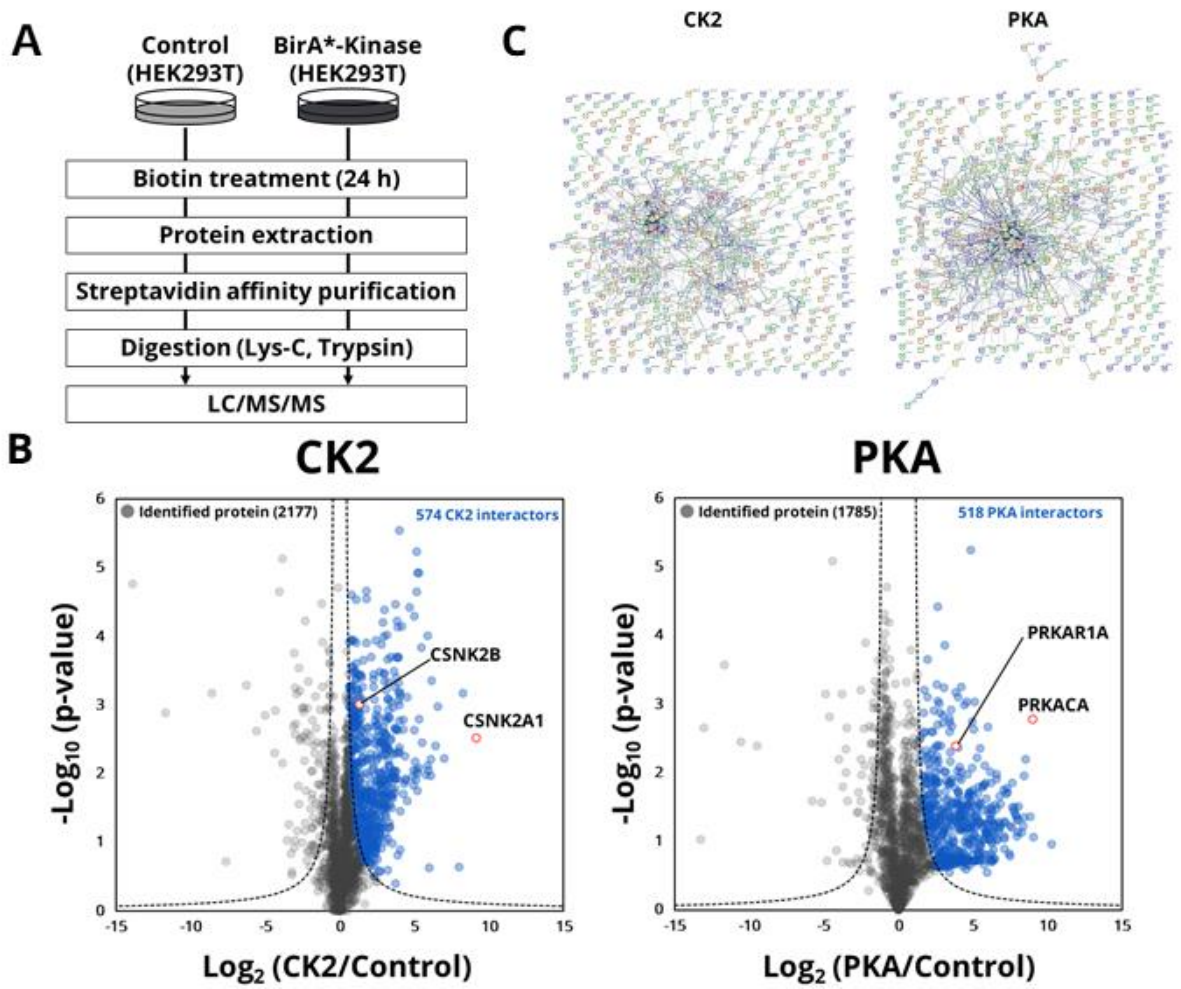

Figure 1. Identification for kinase interacting proteins with BioID.

(A) Workflow to identify kinase interacting proteins with BioID. (B) The ratio of BirA*-kinase to control (log2 (BirA*-kinase/control)) and the negative logged (base 10) *p* values (Welch's t-test) were plotted about each proteins. Proteins were filtered by cutoff lines ($y = c/(x - x_0)$) (c = curvature, x_0 = minimum fold change) as previously described (56). Cutoff parameter for CK2 were $c = 1$ and $x_0 = 0.3$, and for PKA were $c = 1$ and $x_0 = 1$. (C) PPI network of interacting proteins.

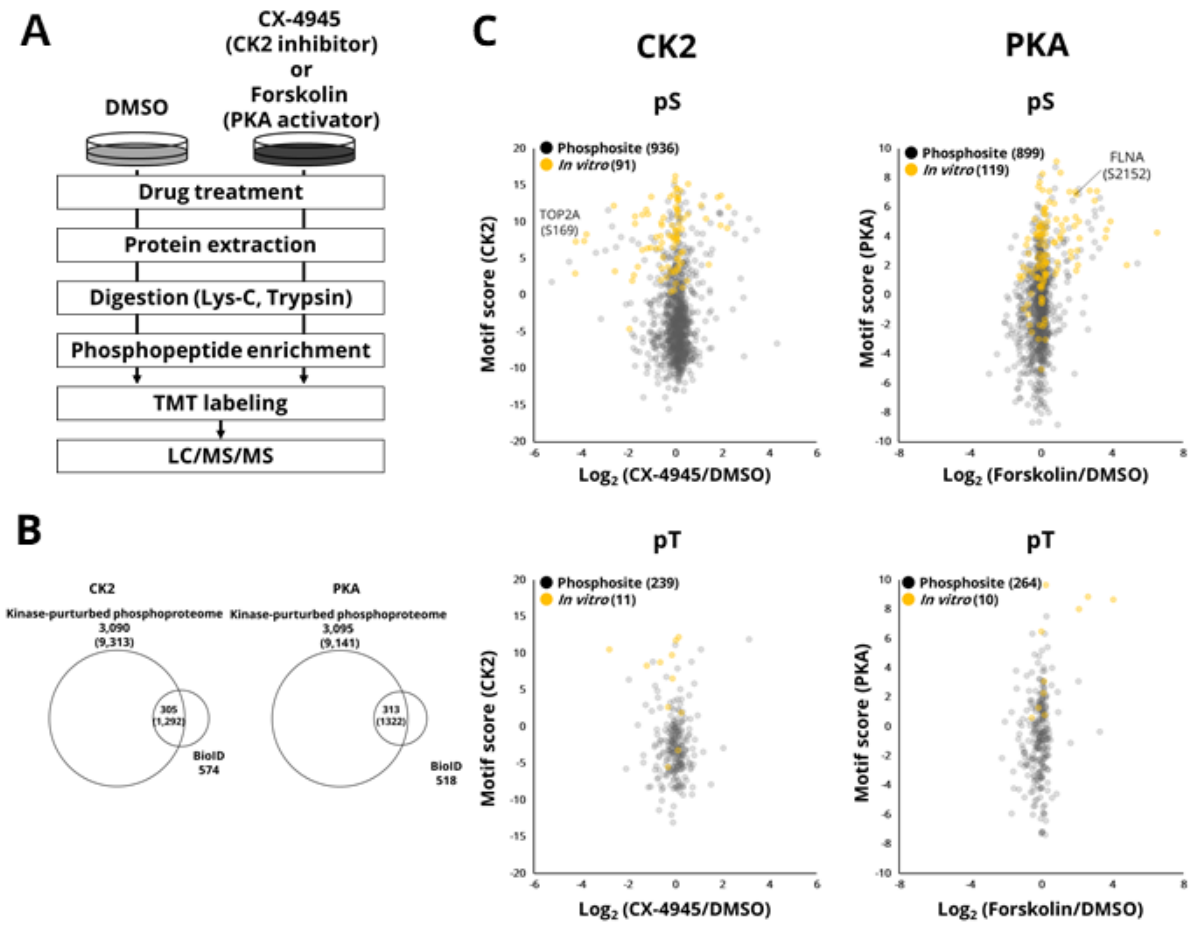

Figure 2. Identification for phosphosites regulated by the target kinase.
(A) Quantitative phosphoproteomics to profile phosphorylation change with the target kinase inhibitor/activator. (B) Overlap of identified proteins between phosphoproteomics and the target kinase interactors. Parentheses means the number of phosphosites. (C) The distribution of phosphorylation ratios and motif scores for phosphosites identified in cells treated with inhibitor or activator. Colors indicate if the sites have been in vitro substrates.

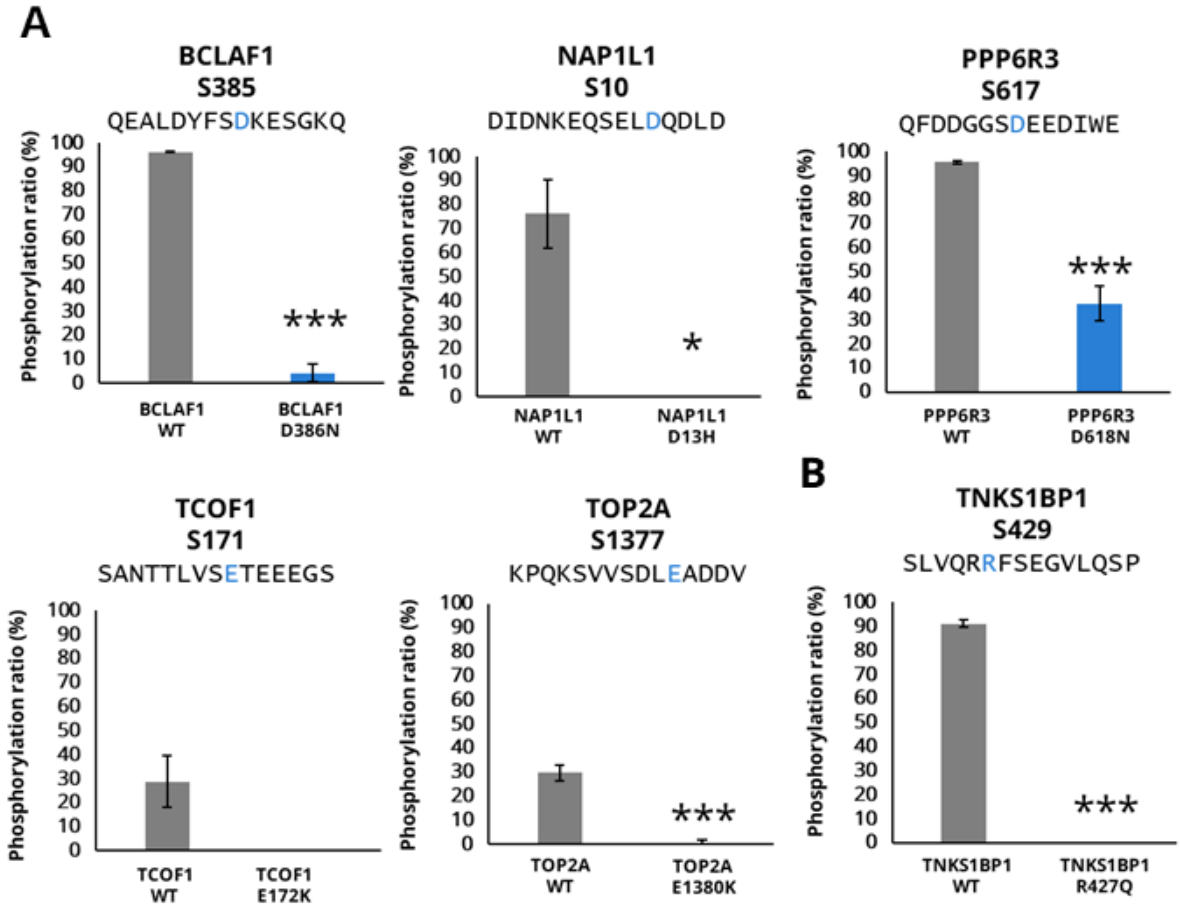

Figure 3. *In vitro* kinase assay against synthetic peptides.
Color indicates mutant residues. (A) CK2. (B) PKA.