Article

Universal quantitative kinase assay based on diagonal SCX chromatography and stable isotope dimethyl labeling provides high-definition kinase consensus motifs for PKA and human Mps1

Marco L Hennrich, Fabio Marino, Vincent Groenewold, Geert J.P.L. Kops, Shabaz Mohammed, and Albert J.R. Heck

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/pr400074f • Publication Date (Web): 19 Mar 2013

Downloaded from <http://pubs.acs.org> on March 29, 2013

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications
High quality. High impact.

Journal of Proteome Research is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036
Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Universal quantitative kinase assay based on diagonal SCX chromatography and stable isotope dimethyl labeling provides high-definition kinase consensus motifs for PKA and human Mps1

Marco L. Hennrich^{†‡□#}, Fabio Marino^{†‡#}, Vincent Groenewold[§], Geert J.P.L. Kops^{§‡}, Shabaz Mohammed^{†‡*}, Albert J.R. Heck^{†‡*}.

[†]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584CH, Utrecht, The Netherlands.

[‡]Netherlands Proteomics Centre, Padualaan 8, 3584CH, Utrecht, The Netherlands.

[□]Current affiliation: European Molecular Biology Laboratory, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

[§]Department of Medical Oncology and Department of Molecular Cancer Research, University Medical Center, Strateneum 2.231, Universiteitsweg 100, 3584CG Utrecht, The Netherlands.

[#]These authors contributed equally to this work.

^{*}To whom correspondence should be addressed. E-mail: a.j.r.heck@uu.nl or s.mohammed@uu.nl

ABSTRACT

In order to understand cellular signaling a clear understanding of kinase-substrate relationships is essential. Some of these relationships are defined by consensus recognition motifs present in substrates making them amendable for phosphorylation by designated kinases. Here, we explore a method that is based on two sequential steps of strong cation exchange chromatography combined with differential stable isotope labeling, to define kinase consensus motifs with high accuracy. We demonstrate the value of our method by evaluating the motifs of two very distinct kinases; cAMP regulated protein kinase A (PKA) and human monopolar spindle 1 (Mps1) kinase, also known as TTK. PKA is a well-studied basophilic kinase with a relatively well-defined motif and numerous known substrates *in vitro* and *in vivo*. Mps1, a kinase involved in chromosome segregation, has been less well characterized. Its substrate specificity is unclear and here we show that Mps1 is an acidophilic kinase with a striking tendency for phosphorylation of threonines. The final outcomes of our work are high-definition kinase consensus motifs for PKA and Mps1. Our generic method, which makes use of proteolytic cell lysates as source for peptide-substrate libraries, can be implemented for any kinase present in the kinome.

KEYWORDS

Proteomics, protein kinase, Mps1, TTK, PKA, consensus motif, kinase assay, strong cation exchange chromatography, COFRADIC, stable isotope dimethyl labeling.

INTRODUCTION

Protein phosphorylation is one of the most explored and important post translational modifications (PTM). The level of protein phosphorylation in processes such as metabolism, transcription, cell cycle progression, cytoskeletal rearrangement, and apoptosis is tightly controlled by protein kinases and phosphatases¹. Not surprisingly, mutations and dysfunctions of protein kinases play causal roles in major human diseases including cancer, immunodeficiencies and endocrine disorders^{2, 3}. In order to fulfill their critical roles, kinases need to recognize their substrates among a vast pool of candidates with high specificity. Distal docking sites, scaffolding proteins, allosteric regulation, localization in given cellular compartments, and recognition of consensus phosphorylation motifs are some of the underlying principles for substrate recognition. Of these, the substrate consensus motif is one of the most important. It is defined as the N- and C-terminal amino acid sequence stretch adjacent to the potential site of phosphorylation⁴.

In the 1990ies, the identification of *in vitro* kinase substrate consensus motifs was performed by the use of large oriented synthetic peptide libraries in which the phosphorable serine or threonine amino acids were fixed and the surrounding amino acids were varied. The substrates phosphorylated by the kinase were then separated from the bulk of non-phosphorylated peptides using a ferric iminodiacetic acid column (IDA), and peptide sequences were obtained by Edman sequencing⁵. This method suffered from severe drawbacks including lack in selectivity at the sites of secondary importance in the consensus sequence, negligence of other possible interactions further located from the phospho-site, and incomplete Edman sequencing. When solid-phase *in vitro* assays could be implemented some of those issues could be resolved. In these methodologies, peptide libraries were immobilized on membranes and incubated with the

kinase of interest in presence of radiolabeled ATP⁶. Unfortunately, these strategies exhibited artifacts due to the nonspecific binding of radiolabeled ATP on the solid support and non-covalent association of the kinase with less optimal substrates due to their high local concentrations^{7, 8}. A more recent *in vitro* approach makes use of synthetic peptide libraries whereby the kinase assay with radiolabeled ATP is performed in solution followed by immobilization⁹. This contributed to a significant improvement in the prediction of *in vitro* kinase substrate consensus motifs. Emerging methods using high-resolution mass spectrometry (MS)¹⁰⁻¹² for the in-depth analysis of phosphopeptides, further improved the identification of putative kinase targets and thus, yielded to a more extensive comprehension of signaling networks^{13, 14}. Still, discovering putative substrates for kinases is challenging. Partially, this is caused by the fact that MS-based identification of phosphopeptides suffers from the sub-stoichiometric level of phosphorylation on peptides compared to their non-modified counterparts¹⁵. Wide-scale MS-based proteomics approaches have the great advantage to render accessible database-searchable large peptide libraries consisting of natural biological sequences derived from appropriate cellular proteomes. This for instance allowed the determination of protease cleavage site consensus motifs^{16, 17}. A similar MS-based approach for determining kinase consensus motifs¹⁸ also made use of peptide libraries allowing protein kinases to choose among thousands of different peptide sequences, targeting only the likely endogenous chemical and positional preferences. This latter method implemented in its workflow several steps to overcome the notorious problems linked to the use of MS-based techniques on complex peptide libraries as represented by a full proteome digest. Not only strong cation exchange (SCX)¹⁹ was used to fractionate and reduce the complexity of the sample, also a phosphatase induced dephosphorylation was implemented to minimize the presence of endogenous phosphate groups.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Following the *in vitro* kinase reaction, a final TiO₂ enrichment was used to pull down the phosphopeptides prior to LC-MS/MS analysis. Despite the developments brought in by this approach, once more, the trivial issue of unambiguously assessing endogenously phosphopeptides from the kinase substrates was not completely solved. Here, we evaluate an alternative high throughput LC-MS/MS-based proteomics method for the definition of consensus motifs of protein kinases. Our approach involves the use of two stages of SCX chromatography (equivalent to combined fractional diagonal chromatography; COFRADIC²⁰⁻²²) where the separation step is used to distinguish between regular and phosphorylated peptides. Our approach is independent of the choice of peptide mixture and thus one can make a peptide library using any cell lysate and any protease allowing access to diverse libraries of natural peptides. The added flexibility in our workflow will help to increase the chance of generating appropriate substrates and improve the chance of identifying the consensus motifs of structurally very diverse protein kinases, but also to enquire those whose specificity is still unknown. Our strategy does not require phosphatase treatment, since it can distinguish between originally present and newly formed phosphorylation sites. We demonstrate the performance of our approach by first determining the kinase motif of the quite well-characterized cAMP regulated protein kinase A (PKA). Subsequently, we define the preferred substrate motif for phosphorylation by human Mps1 kinase, an essential regulator of chromosome segregation.

EXPERIMENTAL SECTION:

Materials:

Complete Mini EDTA-free protease inhibitor cocktail and phosStop phosphatase inhibitor cocktail were obtained from Roche Diagnostic (Germany), metallo-endopeptidase Lys-N from Seikagaku Corporation (Japan), lysyl endopeptidase (Lys-C) from Wako (Richmond, VA) and trypsin endopeptidase from Promega (The Netherlands). Iodacetamide, DL-dithiothreitol (DTT), formaldehyde (37% in water), formaldehyde $^{13}\text{CD}_2$ (20 % in D_2O) and sodium cyanoborodeuterate were purchased from Sigma Aldrich (Germany). Urea, potassium dihydrogen phosphate (KH_2PO_4), potassium chloride (KCl) and formic acid (FA) were obtained from Merck KGaA (Germany). Ammonium bicarbonate (AMBIC) and sodium cyanoborohydrate were obtained from Fluka Analytical (Germany), while HPLC-S grade acetonitrile (ACN) was obtained from Biosolve BV (The Netherlands). Bradford protein assay was purchased from BioRad (Germany), high purity water was obtained from a Millipore Milli-Q system (Billerica, MA) and deuterated formaldehyde CD_2O (20% solution in D_2O) from Isotec (Miamisburg, OH).

Lys-N library preparation for PKA:

HEK293 cells were re-suspended in lysis buffer composed of 8M urea, 50 mM AMBIC pH 8, 1 tablet of PhosStop phosphatase inhibitors, and 1 tablet of cOmplete Mini EDTA-free protease inhibitor cocktail. The cell lysate was sonicated 3 times on ice. Subsequently, centrifugation at 20000 g at 4°C for 20 minutes separated the soluble from the insoluble protein fraction. The soluble fraction was collected and the protein concentration of the lysate was determined by a Bradford protein assay. Proteins were reduced with 2 mM DTT at 56°C for 25 minutes, followed by alkylation with iodacetamide (4 mM) at room temperature for 30 minutes in the dark. An

incomplete digestion was carried out for 4 hours at 37°C using a substrate to Lys-N ratio of 100:1 (w/w)²³. After digestion the sample was acidified with 10% FA, then desalted using Sep-Pak 1CC 50 mg C18 cartridges (Waters Corporation, Milford, MA), and dried down. Dried peptides were re-suspended in 10% FA before the SCX separation.

Trypsin library preparation for Mps1:

Hela S3 cells were digested as described for the PKA experiment with a few modifications. After the alkylation step the digestion was carried out with a first step of lysyl endopeptidase (Lys-C) for 4 hours at 37°C with a protein to enzyme ratio of 75:1 (w/w). Subsequently, the sample was diluted 4 times with 50 mM AMBIC to a urea concentration of 2 M. The second step of digestion was performed with trypsin overnight at 37°C with a substrate to enzyme ratio of 100:1 (w/w). The resulting solution was acidified with 10% FA to quench the reaction, desalted on a Sep-Pak 1CC 50 mg C18 cartridge and the eluate was dried down. Dried peptides were re-suspended in 10% FA before the SCX separation.

SCX separation before the kinase assay:

All SCX experiments were performed on an Agilent 1100/1200 HPLC system (Agilent Technologies, Germany). Peptides from the digests corresponding to 1 mg of protein material for the PKA experiment and approx. 3 mg for the Mps1 experiment were loaded onto a C18 trap column (strata-x 33 µm Polymeric Reversed Phase, 50x4.6 mm, Phenomenex, The Netherlands) for 5 minutes at 300 µL/min using aqueous 0.05% FA as solvent. Subsequently, peptides were eluted for 5 minutes from the trapping column with 80% ACN containing 0.05% FA onto a PolySULFOETHYL A column 200 × 2.1 mm, 5 µm particles, and 200 Å pore size (PolyLC Inc., Columbia, MD) at the same flow rate. Separation was performed using a nonlinear 65 min.

gradient: isocratic for 2 min. at 100% solvent A (5 mM KH_2PO_4 , 30% ACN and 0.05% FA, pH 2.7); from 2 to 10 min. at 3% solvent B (5 mM KH_2PO_4 , 30% ACN, 350 mM KCl and 0.05% FA at pH 2.7); from 10 to 40 min. a gradient to 35% solvent B, and from 40 to 45 min. to 100% solvent B. The column was subsequently washed for 10 min. with solvent B and finally equilibrated with 100% solvent A. Fractions were collected in 1 min. intervals for the first 40 min. and for 3 min. intervals in the last 15 minutes. After evaporation of the solvents, fractionated peptides were resuspended in both cases in kinase buffer as described below.

Protein expression and purification of PKA:

The cAMP-dependent PKA catalytic subunit (New England Biolabs) had a concentration of 2500000 U/mL and an activity of 5000000 U/mg. It was supplied in 50 mM NaCl, 20 mM Tris-HCl at pH 7.5, 1 mM Na_2EDTA , 2 mM DTT and 50% of glycerol.

Protein expression and purification of Mps1:

P2 baculovirus (Invitrogen) was used to facilitate direct transfer of the gene of interest *in vitro* into SF9 cells. pFASTBAC plasmid was used to generate GST-Mps1-wild-type (WT) and – kinase-dead (KD: D664A). The enzymes were then purified using glutathion beads. Detailed procedures are described in²⁴.

In vitro PKA phosphorylation:

Selected fractions from the initial SCX separation were pooled together in order to form 6 samples containing peptides of net charge from +2 to +7. These pools were de-salted using Sep-Pak cartridges followed by drying down and re-suspension in 45 μL of 100 mM Tris-HCl and 20 mM MgCl_2 . The 6 samples were split in two, one half would be used for the *in vitro* assay and

the other half as a control for the assay. 25 μL of kinase reaction buffer composed of 10 mM Tris-HCL, pH 7.5, 20 mM MgCl_2 and 1 μL of the PKA solution containing 2500 U of kinase were added to both aliquots, while 20 mM ATP was added only to the *in vitro* aliquots, whereas no ATP was added to the control aliquots. All the samples were incubated at 37°C for 10 minutes and subsequently, the reaction was quenched with 10% FA.

***In vitro* Mps1 phosphorylation:**

Selected fractions from the initial SCX separation were selected and pooled together in order to form 3 samples containing peptides of net charge from +2 to +4. The 3 pools were de-salted on a Sep-Pak cartridge, dried down and then re-suspended in 30 μL of 500 mM Tris-HCl, 30 μL of 100 mM MgCl_2 , 30 μL of 10 mM DTT and 180 μL of Milli-Q water. The reconstituted samples were split into three. A third of the sample was used for the *in vitro* assay with wild-type Mps1, a second third with the Mps1 kinase-dead mutant and the last aliquot with wild-type Mps1 without addition of ATP. Essentially, the two latter aliquots are expected to contain no phosphorylation activity and are controls. The kinase reaction was performed by adding 10 μL of 2 mM ATP (or 10 μL of Milli-Q water for the control without ATP), 1 μL of wild-type kinase solution to one sample, 1 μL of kinase-dead solution to another sample and 1 μL of wild-type kinase solution to the sample without ATP. The aliquots were incubated for 30 minutes at 37°C and then quenched with 10% formic acid.

On column stable isotope dimethyl labeling:

The on column dimethyl labeling of the samples was performed as described previously²⁵. The PKA sample was labeled with the light dimethyl label and the control with the intermediate label. For the Mps1 experiment the samples containing wild-type kinase, wild-type kinase without ATP

and the kinase-dead mutant were labeled with the heavy, intermediate and light label, respectively. The eluates of the PKA and Mps1 samples were combined according to the SCX pool of origin at a 1:1 and 1:1:1 ratio, respectively.

SCX separation following the *in vitro* kinase assays:

In both experiments each combined labeled sample was subjected to SCX separation as already explained for the first SCX experiment. Newly formed phosphopeptides would elute earlier, due to their decrease in net charge, than peptides that were not phosphorylated by the kinases. Selected fractions were desalted, dried down and then resuspended in 10% FA.

LC-MS/MS analysis:

Selected fractions from the second SCX separation were subjected to nanoLC-MS/MS analysis using an LTQ Orbitrap Velos Mass Spectrometer equipped with an electron transfer dissociation (ETD) source (Thermo Fisher, Germany). Mobile phase buffers for LC separation consisted of 0.1 M acetic acid in water (buffer A) and 0.1 M acetic acid in 80% (buffer B). For the Mps1 experiment trapping and washing of the sample was performed at 5 μ L/min for 10 min with 100% buffer A. The gradient for the separation was from 10 to 25% of buffer B within 107 min. and from 25 to 50% in the following 35 minutes. After the gradient, the column was washed by increasing the buffer B concentration to 100% followed by conditioning the system with 100% buffer A for at least 15 minutes. For the PKA samples a total analysis time of 90 min. was chosen. The gradient ranged from 0 to 30% buffer B in 61 min., then for 3 min. from 30% to 100% buffer B, followed by washing and conditioning the column. The flow rate over the column was kept at 100 nL/min and the column effluent was directly introduced into the ESI source of the MS using an in-house pulled fused silica emitter gold-coated, biased to 1.7 kV. The

mass spectrometer was operated in positive ion mode, from 350 to 1500 m/z in MS mode. We carried out the analysis with a data-dependent mode with MS operating at a resolution of 30,000. MS/MS dependent data were acquired with a Top5 method, where each precursor was fragmented with both HCD and ETD (a total of 10 MS/MS events). Charge state screening was enabled and precursors with unknown charge state or a charge state of 1 were excluded. The normalized collision energy was set to 40% for HCD. The resulting fragments were detected with a resolution of 7500 in the Orbitrap. Parent ions were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap collision cell by ETD and supplemental activation was enabled. The ETD reagent target value was set to 200,000 and the reaction time to 50 ms. The resulting fragments were detected in the LTQ and dynamic exclusion was enabled (exclusion size list 500, exclusion duration 60 s).

Data Analysis:

Each raw data file was processed and quantified by Proteome Discoverer (version 1.3, Thermo Scientific). Top N Peaks filter was selected, where the 10 most abundant peaks in a mass window of 100 Da alongside a signal-to-noise threshold of 1.5 were parsed. ETD non-fragment filter was also implemented with the precursor peak removal asset to 4 Da, charge-reduced precursor removal to 2 Da, and removal of known neutral losses from the charge-reduced precursor to 2 Da and a maximum neutral loss mass set to 120 Da. Only high resolution (HCD) MS/MS spectra were charge de-convoluted with default options using a node of Proteome Discover called "MS2 Spectrum Processor". All generated peak lists were searched using Mascot software (version 2.3.02 Matrix Science). Both Mps1 and PKA data were searched against the Uniprot-Human database. The database search was performed with the following parameters: a mass tolerance of ± 50 ppm for precursor masses; ± 0.8 Da for ETD-ion trap fragment ions; ± 0.08 Da for HCD-

Orbitrap fragment ions, allowing two missed cleavages, and cysteine carbamidomethylation as fixed modification. Light, intermediate and heavy dimethylation were set as a variable modification on peptide N-termini and lysine residues in the case of the Mps1 experiment and only light and intermediate dimethylation for PKA. Methionine oxidation, phosphorylation on serine, threonine and tyrosine were set as variable modifications. The enzyme was specified as Lys-N (metalloendopeptidase) in the case of PKA and trypsin for Mps1. The fragment ion type was specified as electrospray ionization ETD-TRAP, and ESI-QUAD-TOF. The scoring of phosphorylation sites of the identified phosphopeptides was performed by the phosphoRS algorithm (version 2.0)²⁶ implemented in Proteome Discoverer. The dimethyl-based quantitation method was chosen in Proteome Discoverer, with a mass precision requirement of 2 ppm for consecutive precursor measurements. We also allowed spectra with 1 missing channel to be quantified in the case of PKA and 2 missing channels for Mps1. We enabled the options for replacing the missing quantification values with minimum intensity which resulted to be fundamental in our experiment. Percolator²⁷ calculated the target FDR with a strict cut-off of 0.01. After identification and quantification, we combined all results and filtered them with the following criteria: only the PSMs with a site localization probability for the phospho-group of at least 0.75 were selected, Mascot ion score of at least 20, maximum peptide rank 1, high peptide confidence, lowest peptide length 7 and highest unlimited, labeling ratio of at least <0.2 intermediate/light (I/L) in the case of PKA and ratio <0.2 light/heavy and intermediate/heavy (L/H and I/H) for Mps1. If multiple PSMs showed the same sequence then only the one with higher Mascot ion score and pRS site localization probability where selected.

Motif Analysis:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

IceLogo²⁸ was used to obtain the most probable consensus motif of phosphopeptides identified in the *in vitro* kinase assays. Briefly, IceLogo calculates the chance of occurrence (p) of every amino acid in every position in the experimental set. Our experimental set of phosphopeptides was created aligning those at their common phosphorylation site. The reference set for calculating the occurrence was the Uniprot complete proteome set. IceLogo uses statistics to find over- and under-represented amino acids of the experimental dataset compared to the background database in each position flanking the phosphorylation-site. A confidence interval can be set using specific p-values, which in our case was set to 0.001 in order to have a high certainty for the determination of the most likely phosphorylation motifs of the enzymes. In addition, IceLogo was used for visualizing manually selected subsets of peptides.

RESULTS AND DISCUSSION

Current proteomics strategies which exploit peptide libraries originating from natural sources for finding consensus motifs of protein kinases often still face interferences from endogenously phosphorylated peptides. In the here presented strategy, we diminish this issue, which enables us to define *in vitro* kinase motifs with high precision. Unique to our approach is the controlled generation of peptide libraries from natural sources by using an appropriate choice of protease for the kinase being studied. Therefore, our approach is suitable to study kinases with unknown specificity by combining peptide libraries, generated by using a range of proteases, enabling the coverage of more proteome space²⁹. The created peptide libraries were first separated based on their net charge by SCX chromatography at pH 3. Importantly, at pH 3 carboxyl groups are predominantly protonated and have only a minor influence on the net charge, whereas phospho-sites on the peptides are still negatively charged, thus reducing the net charge of peptides by one in comparison to their unphosphorylated counterparts. Thus, the difference in net charge state enables the separation of phosphopeptides from their non-phosphorylated counterparts³⁰⁻³². Libraries of peptides for the *in vitro* kinase assays were selected from the SCX fractions, discarding the SCX fractions that in our set-up contain mainly phosphopeptides. Fractions carrying peptides with the same net charge were pooled together and subsequently split in equal parts for *in vitro* treatment with the kinase and for the control. Prior to mixing the peptides from the kinase assays and controls, they were differentially isotopically dimethyl labeled. The labeling represents one of the key-steps in our workflow enabling to distinguish unambiguously kinase targets from endogenously phosphorylated peptides. After the kinase assay is performed, a second round of SCX chromatography is performed. Newly formed phosphopeptides will have a lower net charge, and will thus elute earlier. This enables the separation from the bulk of

1
2
3 peptides. Phosphopeptides that were originally present will not have their net charge changed
4
5 and so will also elute with the bulk of peptides. Thus, the second SCX separation and the
6
7 dimethyl labeling will allow us to clearly identify newly formed phosphopeptides. Selected SCX
8
9 fractions were then analyzed by LC-MS/MS. The resulting data were then filtered for several
10
11 criteria including confidence of site assignment and observed ratio. Our exact criteria can be
12
13 found in the material and methods section. In order to identify enrichment of selected amino
14
15 acids at positions surrounding the phosphorylation site, IceLogo was used as a predictive tool,
16
17 where the data-set of phosphopeptides was compared to a background dataset based on the whole
18
19 proteome (Figure 1).
20
21
22
23

24
25 ***In vitro* kinase substrate assay for the cAMP regulated protein kinase A**
26

27
28 In an effort to validate our approach we studied cAMP regulated protein kinase A (PKA). It is a
29
30 well-known lysine/arginine directed serine/threonine kinase. Over 160 physiological PKA
31
32 substrates have already been described mostly confirmed by thorough biochemical assays and by
33
34 using strict criteria of eligibility, e.g. the phosphorylation of the targeted protein was confirmed
35
36 both *in vitro* and *in vivo*. Despite the already high number of substrates described, it has been
37
38 hypothesized that these represent only 1% of the total pool of substrates of PKA^{33, 34}.
39
40
41
42 Nevertheless, PKA represents a kinase that is difficult to target by *in vitro* kinase assays and MS-
43
44 based proteomics for a number of reasons. PKA targets basophilic substrates rich in basic
45
46 residues, a challenge for SCX separations. Furthermore, trypsin is not an ideal enzyme to create a
47
48 peptide library which contains substrates of PKA, since trypsin cleaves specifically at lysine or
49
50 arginine, which are part of the predicted phospho-site motif of PKA. We argued that using
51
52 trypsin would eliminate a vast number of potential substrate peptides. The need of a proteolytic
53
54 enzyme able to generate PKA suitable libraries containing relatively long and more basic
55
56
57
58
59
60

1
2
3 peptides amenable for LC-MS/MS analysis, directed our attention towards Lys-C or Lys-N. We
4
5 chose Lys-N due to its cleavage specificity at the N-terminus of lysine, which will lead to a
6
7 slightly higher number of potential substrates when compared to Lys-C. Furthermore, we
8
9 performed a sub-optimal digestion allowing miss-cleavage formation and increasing the
10
11 propensity of RKXS(T), KKXS(T) etc. sequences. The generated proteolytic peptides were
12
13 subsequently separated by SCX, where only the late fractions were further used, since they
14
15 contain mainly highly basic peptides. Fractions, that would contain the same net charge, were
16
17 pooled together. Prior to the *in vitro* kinase assay, the activity of PKA was tested with its proto-
18
19 typical substrate kemptide (data not shown). The pools were split in two, half for the control
20
21 reaction without the addition of ATP, and half for the *in vitro* kinase assay. After the assay, the
22
23 peptide pools were intermediate and light dimethyl labeled for the control and for the assayed
24
25 aliquots, respectively. Post incubation and labeling, the pools were appropriately mixed together
26
27 again and concentrated in a vacuum centrifuge. The reconstituted peptide pools were once again
28
29 fractionated by SCX and subjected to analysis by LC-MS/MS. Considering we were expecting
30
31 our newly formed phosphopeptides to be highly basic and highly charged, we opted for the use
32
33 of both ETD and HCD for sequencing¹² (see Figure 2 for outline of workflow). As expected, we
34
35 identified the majority of newly formed phosphopeptides from SCX pools corresponding to a
36
37 higher net charge i.e. multiple basic residues. Moreover, almost all identified putative PKA
38
39 substrate peptides were identified by using ETD. An example of MS and MS/MS spectra of a
40
41 putative PKA target is shown in Figure 3. From the analysis we could quantify over 2100
42
43 peptides (Supplementary Table 1) and more than 150 phosphopeptides having a phosphoRS site
44
45 probability greater than 75% (Supplementary Table 2). After filtering for a cutoff ratio
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(Light/Medium < 0.2, Supplementary Figure 1), we identified more than 120 putative targets of PKA in our *in vitro* kinase assay (Supplementary Table 3).

The resulting dataset of phosphorylated kinase substrates was evaluated with IceLogo to determine the enrichment of specific amino acids at positions surrounding the phospho-site. The overall motif of PKA detected by IceLogo, considering as position 0 the phospho-site, showed a very high enrichment in arginine at positions -2 and/or -3, while at +1 position hydrophobic amino acid such as I, L and V were over-represented (Figure 4). These data are in very good agreement with known literature, whereby the predominant motif found in *in vivo* studies is RRXS/TX also with a preference of hydrophobic amino acids at position +1³⁴. Furthermore, due to the sensitivity and selectivity of our method, we could investigate in-depth and confirm the positional preference of PKA. We noticed that only 5% of all our *in vitro* targets were phosphorylated at threonine with 95% at serine. Indeed, PKA has also been predicted to have, both *in vivo* and *in vitro*, a marked preference for serine³³. The selectivity of PKA was shown to be XRXS(T)X > RXXS(T)X > RRXS(T)X. Other motifs were also found in our data-set, albeit with much lower frequency, such as KRXS(T)X, RKXS(T)X, RQXS(T)X and others. Also several of these motifs have already been described as bona fide *in vivo* PKA targets^{33, 34}. The observation of motifs such as RKXS(T)X and KKXS(T)X in our dataset confirmed that the choice of Lys-N and the digestion conditions chosen were quite optimal for generating the PKA peptide library.

***In vitro* kinase substrate assay for Mps1, a kinase involved in mitotic spindle orientation**

In order to demonstrate the wide applicability of our approach, we next sought to define sequence preference for phosphorylation by a poorly characterized but essential kinase, human

Mps1, also known as TKK. Mps1 ensures correct chromosome segregation by establishing proper chromosome attachments to the mitotic spindle and by activating the mitotic checkpoint that delays cell division when such attachments have not been established^{35, 36}. Compared to PKA, human Mps1 is less well characterized. Some genuine Mps1 substrates have been identified, including the conserved centromeric protein KNL1³⁶. In addition, Mps1 auto-phosphorylates in cis and trans^{24, 37, 38}. Still, a high-definition prediction of the authentic Mps1 phosphorylation motif is missing. Mps1 is thought to phosphorylate substrate peptides enriched in acidic amino acids (E and D)^{39, 40}, which makes it distinct from basophilic kinases like PKA. In order to create suitable peptides examination as Mps1 substrates, we picked trypsin as protease for the library generation. The resulting tryptic peptides were separated by SCX and fractions that would contain peptides of the same net charge were pooled. Subsequently, each pool was split into three aliquots. One third was used for the *in vitro* assay with wild-type Mps1, another third with a kinase-deficient mutant of Mps1 and the remaining part with wild-type Mps1 without addition of ATP. Essentially, the two latter aliquots, the controls, are expected to contain no phosphorylation activity. Light and intermediate dimethyl labels were chosen for the controls and the heavy form for the assayed aliquot. The reconstituted pools were then separated by SCX chromatography again. Subsequently, selected fractions were analyzed by LC-MS/MS. After the data analysis, the putative targets were defined as phosphorylated peptides with the heavy channel being at least 5 times more abundant than the other two channels (controls) (Figure 2). From the analysis we could quantify more than 3500 peptides (Supplementary Table 4) and more than 560 phosphopeptides having a phosphoRS site probability greater than 75% (Supplementary Table 5). After filtering for a ratio greater than five (Heavy/Light > 5 and Medium/Light > 5, Supplementary Figure 2), we gathered more than 500 potential substrates

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(Supplementary Table 6). Once again, we used the IceLogo program to identify amino acids that appeared to be enriched in the putative targets of Mps1. From this data, we could derive a putative consensus sequence motif for human Mps1, considering as position 0 the phospho-site. This motif contains a high enrichment in acidic amino acids in the positions -2, and/or -3 with a preference at -2, while in the +3 position a predominant preference for the hydrophobic branched-chain amino acids leucine, valine and isoleucine is observed (Figure 5). We noticed that over 90% of the peptide targets were phosphorylated on a threonine, and only a few on a serine. This clear preference for threonine is considerable and highly selective, since threonine phosphorylations are less common in human⁴¹. The few studies reporting on the substrate repertoire of Mps1 are consistent with these preferences: the target sequences in the budding yeast Mps1p substrates Spc29 and Spc110 are phosphorylated mostly on threonines and have a predominance of acidic residues, most commonly at position -2^{42, 43}. Similarly, phosphorylation of human Borealin as well as yeast and human KNL1/Spc105 by Mps1 occurs exclusively on threonines and particularly on those preceded by acidic residues⁴⁴⁻⁴⁷. Moreover, the related auto-phosphorylation sites of Mps1^{24, 37-39} also have an E/D/N/Q preference at -2^{40, 42, 48-50}. However, although a bias for I/V/L at position +3 can be seen in the auto-phosphorylated sequences, no such bias can be detected in KNL1/Spc105. Some of the numerous MELT-like motifs in this protein are bona fide Mps1 substrate sequences but the +3 position is highly variable. Mps1 may therefore select substrates based mostly on threonine preceded by acidic residues. Our description of preferred Mps1 substrate sequences will greatly aid identification of bona fide functional Mps1 targets, and as such will help define the molecular regulatory network that guards against chromosome segregation errors.

CONCLUSIONS

We present a sensitive method designed for evaluating kinase consensus motifs with high accuracy. In this method, we create peptide libraries originating from the same organism as the studied kinase and combine it with a COFRADIC approach based on SCX chromatography. In order to make this strategy suitable to large peptide libraries, we combined it with stable isotope labeling. We evidenced its merits by finely discerning and thus, identifying several *in vitro* targets of two totally different protein kinases, namely the basophilic PKA and the acidophilic Mps1. Furthermore, from the statistical analysis of the two datasets, we could confirm the already well defined kinase consensus motif of human PKA and retrieved a high accuracy kinase consensus motif for human Mps1. Therefore, the use of our flexible and generic method can be extended potentially to any kind of kinase existing in the kinome.

ACKNOWLEDGEMENTS

The authors acknowledge all members of the Heck-group. This research was performed within the framework of the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program. Additionally, the Netherlands Proteomics Centre, a program embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support as well as the Netherlands Organization for Scientific Research (NWO) with the VIDI grant for SM (700.10.429).

SUPPORTING INFORMATION AVAILABLE

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary Figure 1 and 2 as well as Table 1 to 6 are available via the Internet at
<http://pubs.acs.org>.

REFERENCES:

1. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002**, 298, (5600), 1912-34.
2. Blume-Jensen, P.; Hunter, T., Oncogenic kinase signalling. *Nature* **2001**, 411, (6835), 355-65.
3. Stenberg, K. A.; Riikonen, P. T.; Vihinen, M., KinMutBase, a database of human disease-causing protein kinase mutations. *Nucleic Acids Res* **1999**, 27, (1), 362-4.
4. Ubersax, J. A.; Ferrell, J. E., Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **2007**, 8, (7), 530-41.
5. Songyang, Z.; Blechner, S.; Hoagland, N.; Hoekstra, M. F.; Piwnicka-Worms, H.; Cantley, L. C., Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr Biol* **1994**, 4, (11), 973-82.
6. Rodriguez, M.; Li, S. S.; Harper, J. W.; Songyang, Z., An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J Biol Chem* **2004**, 279, (10), 8802-7.
7. Yaffe, M. B., Novel at the library. *Nat Methods* **2004**, 1, (1), 13-4.
8. Manning, B. D.; Cantley, L. C., Hitting the target: emerging technologies in the search for kinase substrates. *Sci STKE* **2002**, 2002, (162), pe49.
9. Hutti, J. E.; Jarrell, E. T.; Chang, J. D.; Abbott, D. W.; Storz, P.; Toker, A.; Cantley, L. C.; Turk, B. E., A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods* **2004**, 1, (1), 27-9.
10. Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **2007**, 6, (11), 1942-51.
11. Wiesner, J.; Premisler, T.; Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008**, 8, (21), 4466-83.
12. Frese, C. K.; Altelaar, A. F.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res* **2011**, 10, (5), 2377-88.
13. Xue, L.; Wang, W. H.; Iliuk, A.; Hu, L.; Galan, J. A.; Yu, S.; Hans, M.; Geahlen, R. L.; Tao, W. A., Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc Natl Acad Sci U S A* **2012**, 109, (15), 5615-20.
14. Singh, S. A.; Winter, D.; Bilimoria, P. M.; Bonni, A.; Steen, H.; Steen, J. A., FLEXIQinase, a mass spectrometry-based assay, to unveil multikinase mechanisms. *Nat Methods* **2012**, 9, (5), 504-8.
15. Zhou, H.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J.; Mohammed, S., Towards a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res* **2012**.
16. Schilling, O.; Overall, C. M., Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **2008**, 26, (6), 685-94.
17. Schilling, O.; auf dem Keller, U.; Overall, C. M., Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. *Methods Mol Biol* **2011**, 753, 257-72.
18. Kettenbach, A. N.; Wang, T.; Faherty, B. K.; Madden, D. R.; Knapp, S.; Bailey-Kellogg, C.; Gerber, S. A., Rapid determination of multiple linear kinase substrate motifs by mass spectrometry. *Chem Biol* **2012**, 19, (5), 608-18.
19. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **2004**, 101, (33), 12130-5.
20. Gevaert, K.; Staes, A.; Van Damme, J.; De Groot, S.; Hugelier, K.; Demol, H.; Martens, L.; Goethals, M.; Vandekerckhove, J., Global phosphoproteome analysis on human HepG2 hepatocytes using reversed-phase diagonal LC. *Proteomics* **2005**, 5, (14), 3589-99.

21. Staes, A.; Impens, F.; Van Damme, P.; Ruttens, B.; Goethals, M.; Demol, H.; Timmerman, E.; Vandekerckhove, J.; Gevaert, K., Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* **2011**, 6, (8), 1130-41.
22. Staes, A.; Van Damme, P.; Helsens, K.; Demol, H.; Vandekerckhove, J.; Gevaert, K., Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **2008**, 8, (7), 1362-70.
23. Taouatas, N.; Heck, A. J.; Mohammed, S., Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions. *J Proteome Res* **2010**, 9, (8), 4282-8.
24. Jelluma, N.; Brenkman, A. B.; McLeod, I.; Yates, J. R.; Cleveland, D. W.; Medema, R. H.; Kops, G. J., Chromosomal instability by inefficient Mps1 auto-activation due to a weakened mitotic checkpoint and lagging chromosomes. *PLoS One* **2008**, 3, (6), e2415.
25. Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **2009**, 4, (4), 484-494.
26. Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K., Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* **2011**, 10, (12), 5354-62.
27. Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **2007**, 4, (11), 923-5.
28. Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K., Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **2009**, 6, (11), 786-7.
29. Peng, M.; Taouatas, N.; Cappadona, S.; van Breukelen, B.; Mohammed, S.; Scholten, A.; Heck, A. J., Protease bias in absolute protein quantitation. *Nat Methods* **2012**, 9, (6), 524-5.
30. Di Palma, S.; Hennrich, M. L.; Heck, A. J.; Mohammed, S., Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics* **2012**, 75, (13), 3791-813.
31. Mohammed, S.; Heck, A., Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Curr Opin Biotechnol* **2011**, 22, (1), 9-16.
32. Hennrich, M. L.; van den Toorn, H. W.; Groenewold, V.; Heck, A. J.; Mohammed, S., Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. *Anal Chem* **2012**, 84, (4), 1804-8.
33. Shabb, J. B., Physiological substrates of cAMP-dependent protein kinase. *Chem Rev* **2001**, 101, (8), 2381-411.
34. Ruppelt, A.; Tasken, K., Physiological substrates of PKA and PKG. In Oberprieler, N.; Magklaras, G., Eds. Academic Press Inc.: Handbook of Cell Signaling, 2009; Vol. Three-Volume Set 2 ed.
35. Kops, G. J.; Shah, J. V., Connecting up and clearing out: how kinetochore attachment silences the spindle assembly checkpoint. *Chromosoma* **2012**, 121, (5), 509-25.
36. Vleugel, M.; Hoogendoorn, E.; Snel, B.; Kops, G. J., Evolution and function of the mitotic checkpoint. *Dev Cell* **2012**, 23, (2), 239-50.
37. Kang, J.; Chen, Y.; Zhao, Y.; Yu, H., Autophosphorylation-dependent activation of human Mps1 is required for the spindle checkpoint. *Proc Natl Acad Sci U S A* **2007**, 104, (51), 20232-7.
38. Mattison, C. P.; Old, W. M.; Steiner, E.; Huneycutt, B. J.; Resing, K. A.; Ahn, N. G.; Winey, M., Mps1 activation loop autophosphorylation enhances kinase activity. *J Biol Chem* **2007**, 282, (42), 30553-61.
39. Dou, Z.; von Schubert, C.; Körner, R.; Santamaria, A.; Elowe, S.; Nigg, E. A., Quantitative mass spectrometry analysis reveals similar substrate consensus motif for human Mps1 kinase and Plk1. *PLoS One* **2011**, 6, (4), e18793.
40. Mok, J.; Kim, P. M.; Lam, H. Y.; Piccirillo, S.; Zhou, X.; Jeschke, G. R.; Sheridan, D. L.; Parker, S. A.; Desai, V.; Jwa, M.; Cameroni, E.; Niu, H.; Good, M.; Remenyi, A.; Ma, J. L.; Sheu, Y. J.; Sassi, H. E.; Sopko, R.; Chan, C. S.; De Virgilio, C.; Hollingsworth, N. M.; Lim, W. A.; Stern, D. F.; Stillman, B.; Andrews, B. J.;

- Gerstein, M. B.; Snyder, M.; Turk, B. E., Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci Signal* **2010**, 3, (109), ra12.
41. Mann, M.; Ong, S. E.; Grønborg, M.; Steen, H.; Jensen, O. N.; Pandey, A., Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* **2002**, 20, (6), 261-268.
42. Holinger, E. P.; Old, W. M.; Giddings, T. H.; Wong, C.; Yates, J. R.; Winey, M., Budding yeast centrosome duplication requires stabilization of Spc29 via Mps1-mediated phosphorylation. *J Biol Chem* **2009**, 284, (19), 12949-55.
43. Friedman, D. B.; Kern, J. W.; Huneycutt, B. J.; Vinh, D. B.; Crawford, D. K.; Steiner, E.; Scheiltz, D.; Yates, J.; Resing, K. A.; Ahn, N. G.; Winey, M.; Davis, T. N., Yeast Mps1p phosphorylates the spindle pole component Spc110p in the N-terminal domain. *J Biol Chem* **2001**, 276, (21), 17958-67.
44. Jelluma, N.; Brenkman, A. B.; van den Broek, N. J.; Cruijssen, C. W.; van Osch, M. H.; Lens, S. M.; Medema, R. H.; Kops, G. J., Mps1 phosphorylates Borealin to control Aurora B activity and chromosome alignment. *Cell* **2008**, 132, (2), 233-46.
45. London, N.; Ceto, S.; Ranish, J. A.; Biggins, S., Phosphoregulation of Spc105 by Mps1 and PP1 regulates Bub1 localization to kinetochores. *Curr Biol* **2012**, 22, (10), 900-6.
46. Shepperd, L. A.; Meadows, J. C.; Sochaj, A. M.; Lancaster, T. C.; Zou, J.; Buttrick, G. J.; Rappsilber, J.; Hardwick, K. G.; Millar, J. B., Phosphodependent recruitment of Bub1 and Bub3 to Spc7/KNL1 by Mph1 kinase maintains the spindle checkpoint. *Curr Biol* **2012**, 22, (10), 891-9.
47. Yamagishi, Y.; Yang, C. H.; Tanno, Y.; Watanabe, Y., MPS1/Mph1 phosphorylates the kinetochore protein KNL1/Spc7 to recruit SAC components. *Nat Cell Biol* **2012**, 14, (7), 746-52.
48. Kemmler, S.; Stach, M.; Knapp, M.; Ortiz, J.; Pfannstiel, J.; Ruppert, T.; Lechner, J., Mimicking Ndc80 phosphorylation triggers spindle assembly checkpoint signalling. *EMBO J* **2009**, 28, (8), 1099-110.
49. Shimogawa, M. M.; Graczyk, B.; Gardner, M. K.; Francis, S. E.; White, E. A.; Ess, M.; Molk, J. N.; Ruse, C.; Niessen, S.; Yates, J. R.; Muller, E. G.; Bloom, K.; Odde, D. J.; Davis, T. N., Mps1 phosphorylation of Dam1 couples kinetochores to microtubule plus ends at metaphase. *Curr Biol* **2006**, 16, (15), 1489-501.
50. Araki, Y.; Gombos, L.; Migueleti, S. P.; Sivashanmugam, L.; Antony, C.; Schiebel, E., N-terminal regions of Mps1 kinase determine functional bifurcation. *J Cell Biol* **2010**, 189, (1), 41-56.

Figure Legends

Figure 1. Experimental Workflow. Natural peptide libraries optimized for the individual kinases are created using different proteases. The resulting proteolytic peptides are separated by SCX chromatography. Fractions containing peptides with the same net charge are pooled together and then split to create aliquots for the controls and the *in vitro* kinase assay. After the assay the peptides are labeled by stable isotope dimethyl labeling. Corresponding labeled peptides are pooled back together and are separated again by SCX chromatography. Unmodified peptides elute with identical retention times, while the newly phosphorylated peptides will elute earlier due to their decreased net charge after addition of the phosphate group. Selected SCX fractions are then analyzed by nanoLC-MS/MS. Targets of the kinase will show a high ratio or only the channel of the kinase assay. The putative targets are analyzed by IceLogo to retrieve the sequence motif of the kinase.

Figure 2. Workflow for the *in vitro* kinase assays of PKA and Mps1. Lys-N and trypsin are used to create a suitable peptide library for PKA and Mps1, respectively. Peptides are separated by SCX chromatography and the fractions containing mainly peptides with a net charge of +2, +3, +4, +5, +6 and +7 for PKA and +2, +3 and +4 for Mps1 are pooled together. The pools are then split in two aliquots for the control and the kinase assay with PKA and in three aliquots for the two controls and the Mps1 kinase assay. Subsequently, the samples are differentially dimethyl labeled. Light and intermediate labels are used for the PKA kinase assay and the assay without addition of ATP (control), while light, intermediate and heavy isotope labels are used for the kinase-dead Mps1 control experiment, the wild-type Mps1 assay without addition of ATP (control 2) and the wild-type Mps1 kinase assays, respectively. Next, samples were mixed again and separated by SCX, enabling the separation of phosphorylated peptides from their unmodified counterparts. Enriched phosphofractions are analyzed by LC-MS/MS using both ETD and HCD. After database searching and filtering the list of phosphopeptides is processed with IceLogo for deciphering the amino acid distribution on the surrounding positions of the phospho-site, generating a high-precision consensus motif.

Figure 3. MS and MS/MS spectra of an *in vitro* phosphorylated peptide. In A the MS spectrum of the putative PKA target KSPPREGpSQGELTPANSQSR(O)MSTN(O)M is shown. In the red box the light labeled phosphorylated peptide is observed, whereas the corresponding intermediate labeled peptide (in the blue box) is absent. In B the corresponding ETD fragmentation spectrum of the light peptide is displayed. The spectrum contains extensive series of z and c ions from which the peptide and phospho-site can be easily identified.

Figure 4. High precision PKA consensus motifs and statistically relevant sub-motifs. The unambiguously identified 130 PKA target peptides all exhibited a pRS site probability > 0.75 , a Mascot ion score ≥ 20 and a light/intermediate ratio > 5 . These phosphopeptides were converted into aligned sequences with the phosphorylation site at position 8. The enrichment of each amino acid surrounding the phospho-sites was calculated using IceLogo with a p value of 0.001. Over-represented amino acids are displayed on the top of the figure with their relative frequencies (y axe) in each position, while under-represented ones appear at the bottom. In each figure “n” represents how many sequences were used to calculate the statistics with IceLogo. The overall consensus motif is displayed at the top. In the five lower panels sub-motifs are depicted which were manually selected out of the 130 PKA target. These were illustrated with IceLogo.

Figure 5. High precision Mps1 consensus motif. The unambiguously identified over 500 Mps1 target peptides all exhibited a pRS site probability > 0.75 , a Mascot ion score ≥ 20 and a light/heavy and intermediate/heavy ratio > 5 . These phosphopeptides were converted into aligned sequences with the phosphorylation site at position 8. Statistics to find over- and under-represented amino acids of the experimental dataset compared to the background database in each position flanking the phosphorylation-site were calculated using IceLogo with a p value of 0.001. Over-represented amino acids are displayed on the top figure with their relative frequencies (y axe) in each position, while under-represented ones appear on the bottom. In each figure “n” represents how many sequences were used to calculate the statistics with IceLogo. The overall consensus motif is displayed at the top. In the five lower panels sub-motifs are

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

depicted which were manually selected out of the Mps1 target peptides. These were illustrated with IceLogo.

Supplementary Figure 1. Scatter plot of the distribution of phosphopeptides identified in the PKA kinase assay. The plotted over 150 unambiguously identified phosphopeptides (Supplementary Table 2) all exhibited a pRS site probability greater than 75% and a Mascot ion score equal to or greater than 20. The Log₂ ratios (Intermediate/Light) of the phosphopeptides are plotted against the Mascot ion score. Two distinct distributions corresponding to the PKA targets (red dots) and non-targets (blue dots) can be distinguished. The dashed line indicates the applied Mascot ion score threshold of 20. The maximum allowed ratio in the quantification software was 100000 corresponding to a Log₂ of 16.6.

Supplementary Figure 2. Scatter plot of the distribution of phosphopeptides identified in the Mps1 kinase assay. The plotted over 560 unambiguously identified phosphopeptides (Supplementary Table 5) all exhibited a pRS site probability greater than 75% and a Mascot ion score equal to or greater than 20. The Log₂ ratios (Heavy/Light) of the phosphopeptides are plotted against the Mascot ion score. Two distinct distributions corresponding to the Mps1 targets (red dots) and non-targets (blue dots) can be distinguished. The dashed line indicates the applied Mascot ion score threshold of 20. The maximum allowed ratio in the quantification software was 100000 corresponding to a Log₂ of 16.6.

Digestion

Kinase Assay

Control

Unmodified peptides

Stable Isotope Dimethyl Labeling

Motif Analysis

nanoLC-MS

ACS Paragon Plus Environment

1 1st SCX

2
3
4
5
6
7
8
9

10 Kinase Assay +
11 Isotope Labeling
12

13
14
15
16
17
18
19
20
21
22
23
24
25

Page 33 of 33 Journal of Proteome Research

Digestion

Kinase Assay

Control

Unmodified peptides

Stable Isotope Dimethyl Labeling

- 1
- 2
- 3
- 4
- 5
- 6
- 7

Motif Analysis

ACS Paragon Plus Environment

nanoeC-MS