Methods in Molecular Biology 1355

Springer Protocols

Louise von Stechow *Editor*

Phospho-Proteomics

Methods and Protocols

Second Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Phospho-Proteomics

Methods and Protocols

Edited by

Louise von Stechow

Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Humana Press

Editor
Louise von Stechow
Proteomics Program, Novo Nordisk
Foundation Center for Protein Research
Faculty of Health and Medical Sciences,
University of Copenhagen
Copenhagen, Denmark

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3048-7 ISBN 978-1-4939-3049-4 (eBook)
DOI 10.1007/978-1-4939-3049-4

Library of Congress Control Number: 2015945106

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

Phosphorylation of proteins is a crucial mechanism for regulating cell signaling processes and critically involved in essentially all physiological and pathological processes. Phosphorylation-mediated signaling processes are often deregulated in diseases, such as cancer. Monitoring protein phosphorylation in a given cell or tissue is vital for a better understanding of biological mechanisms underlying physiological and pathological processes.

Phosphoproteomics: Methods and Protocols presents protocols and strategies for identification, evaluation, and quantitation of protein phosphorylation on Tyr, Ser, and Thr residues. This volume summarizes protocols focusing on techniques for the specific enrichment of phosphopeptides and phosphoproteins; different labeling strategies for quantitative phosphoproteomics; high-throughput mass spectrometry-based phosphoproteome analyses and phospho flow cytometry; and bioinformatics strategies for phosphoproteomics data analysis and integration. Furthermore, several protocols concentrate on the identification of kinase-substrate relationships by both high- and low-throughput approaches. In addition to detailed protocols, a number of reviews summarize various enrichment strategies as well as bioinformatics tools for phosphosite and kinase motif prediction and highlight the potential use of those methods in various contexts.

Phosphoproteomics: Methods and Protocols is targeted towards scientists who wish to identify the specific role of phosphorylated proteins in a given biological context by providing a great overview of suitable methods currently used in the field. It is written for a broad audience ranging from researchers who are unfamiliar with proteomic techniques to those with more experience in the field. This volume highlights phosphoproteomics methods and technologies that are widely used and have been well established and furthermore features several protocols that pinpoint new directions in the phosphoproteomics field. Those range from the high-throughput predictions of kinase-substrate relationships to strategies for quantitative phosphoproteomics in tissue samples and meaningful interpretation of large-scale phosphoproteomics analyses using bioinformatics tools.

Copenhagen, Denmark ***Louise von Stechow***

Contents

Contributors

Ruth Hogue Angeletti • *Laboratory for Macromolecular Analysis and Proteomics, Albert Einstein College of Medicine, Bronx, NY, USA; Department for Molecular and Developmental Biology, Albert Einstein College of Medicine, Bronx, NY, USA*

Justine V. Arrington • *Department of Chemistry, Purdue University, West Lafayette, IN, USA*

Íris L. Batalha • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*

Tanveer S. Batth • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Erin E. Carlson • *Department of Chemistry, Indiana University, Bloomington, IN, USA; Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, USA; Department of Chemistry, University of Minnesota, Minneapolis, MN, USA*

Lei Cheng • *Department of Biomedicine and Center for Interactions of Proteins in Epithelial Transport, Aarhus University, Aarhus, Denmark*

Helen J. Cooper • *School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK*

Andrew J. Creese • *School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK*

Margherita Dell'Aica • *Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany*

Clarissa Dickhut • *Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany*

Kasper Engholm-Keller • *Children's Medical Research Institute, Westmead, Sydney, Australia; Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark*

Humberto Gonczarowska-Jorge • *Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany; CAPES Foundation, Ministry of Education of Brazil, Brasília–DF, Brazil*

Elizabeth A. Gordon • *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA*

Chao Han • *Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuchang Moshan, Wuhan, China; University of Chinese Academy of Sciences, Beijing, China*

David H. Hawke • *Department of Systems Biology, The University of Texas, MD Anderson Cancer Center, Houston, TX, USA*

Jason D. Hoffert • *Epithelial Systems Biology Laboratory, National Heart, Lung and Blood Institute, NIH, Bethesda, MD, USA*

Lars J. Jensen • *Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Rosa R. Jersie-Christensen • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Eiji Kinoshita • *Department of Functional Molecular Science, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan*

Emiko Kinoshita-Kikuta • *Department of Functional Molecular Science, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan*

Mark A. Knepper • *Epithelial Systems Biology Laboratory, National Heart, Lung and Blood Institute, NIH, Bethesda, MD, USA*

Tohru Koike • *Department of Functional Molecular Science, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan*

Johannes Landskron • *Biotechnology Centre, University of Oslo, Oslo, Norway; K.G. Jebsen Centre for Cancer Immunotherapy, University of Oslo, Oslo, Norway; Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo, Oslo, Norway*

Martin R. Larsen • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark*

Alexander Leitner • *Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland*

Sebastian Lourido • *Whitehead Institute for Biomedical Research, Cambridge, MA, USA*

Boris Macek • *Proteome Center Tuebingen, Interfaculty Institute for Cell Biology, University of Tuebingen, Tuebingen, Germany*

Ivan Mijakovic • *Department of Systems and Synthetic Biology, Chalmers University of Technology, Göteborg, Sweden*

Stephanie Munk • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Ryohei Narumi • *Laboratory for Synthetic Biology, RIKEN Quantitative Biology Center, Osaka, Japan; Laboratory of Proteome Research, National Institute of Biomedical Innovation, Osaka, Japan*

Heinz Nika • *Laboratory for Macromolecular Analysis and Proteomics, Albert Einstein College of Medicine, Bronx, NY, USA; Department for Molecular and Developmental Biology, Albert Einstein College of Medicine, Bronx, NY, USA*

Jesper V Olsen • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Trairak Pisitkun • *Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand*

Vaishnavi Ravikumar • *Proteome Center Tuebingen, Interfaculty Institute for Cell Biology, University of Tuebingen, Tuebingen, Germany*

Jan C. Refsgaard • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Ana Cecília A. Roque • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*

Daniel A. Rothenberg • *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA*

Abida Sultan • *Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

W. Andy Tao • *Department of Biochemistry, Purdue University, West Lafayette, IN, USA; Department of Chemistry, Purdue University, West Lafayette, IN, USA; Department of Medicinal Chemistry & Molecular Pharmacology, Purdue University, West Lafayette, IN, USA; Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN, USA*

Kjetil Taskén • *Biotechnology Centre, University of Oslo, Oslo, Norway; K.G. Jebsen Centre for Cancer Immunotherapy, University of Oslo, Oslo, Norway; Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo, Oslo, Norway; K.G. Jebsen Inflammation Research Centre, University of Oslo, Oslo, Norway; Department of Infectious Diseases, University of Oslo, Oslo, Norway*

Tine E. Thingholm • *Division of Translational Cancer Research, Lund University, Lund, Sweden; Lund Stem Cell Center, Lund University, Lund, Sweden*

Takeshi Tomonaga • *Laboratory of Proteome Research, National Institute of Biomedical Innovation, Health and Nutrition, Osaka, Japan*

Forest M. White • *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA*

Kaelyn E. Wilke • *Department of Chemistry, Indiana University, Bloomington, IN, USA*

Liang Xue • *Department of Biochemistry, Purdue University, West Lafayette, IN, USA*

Pingfang Yang • *Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuchang Moshan, Wuhan, China*

René P. Zahedi • *Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany*

Hongyan Zhao • *School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK*

Part I

Techniques for Tagging Phospho-Moieties

Chapter 1

Thiol-ene-Enabled Detection of Thiophosphorylation as a Labeling Strategy for Phosphoproteins

Kaelyn E. Wilke and Erin E. Carlson

Abstract

The adenosine triphosphate (ATP) analogue adenosine 5′-O-(3-thiotriphosphate) (ATPγS) has been applied as a tool to study kinase-substrate phosphorylation. Not only does the transfer of a thiophosphate group represent a unique modification amid the phosphoproteome, but it can also be stable to phosphatase activity. However, detection of this species is complicated due to the similar chemical reactivity of thiophosphate and proteinaceous thiols. Here, we describe a novel method for detection of protein thiophosphorylation utilizing the thiol-ene reaction. By first chemoselectively capping protein thiols through radical chemistry, kinase-catalyzed thiophosphorylation can be visualized specifically.

Key words Kinase, Substrate, Phosphoprotein, Thiophosphorylation, Thiol-ene, Thiol capping, ATPγS

1 Introduction

Kinase-mediated thiophosphorylation has been established as an alternative method for studying phosphoproteins where adenosine 5′-O-(3-thiotriphosphate) (ATPγS) is used in place of the native co-substrate, adenosine triphosphate (ATP). With use of this molecule, kinases catalyze the transfer of a thiophosphoryl group to their substrates [1–3]. Because ATPγS is a non-endogenous nucleotide, its turnover yields new, atypically modified substrates among the vast phosphoproteome that includes steady-state phosphorylated proteins [4]. Using mutant kinases that accept modified ATPγS analogues, this strategy has been successful for the identification of direct kinase-substrate pairs in complex milieus [1, 2, 5–7]. Additionally, thiophosphorylated proteins are resistant to phosphatase activity, conferring added stability to this covalent modification [8, 9]. Thus, the thiophosphate provides a handle for identification of posttranslationally modified biomolecules.

Thiophosphorylated substrates and the copious cysteine nucleophiles in protein mixtures share similar reactivities, making it challenging to specifically isolate thiophosphorylated species.

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_1,

Efforts have been made to improve strategies to differentiate thiophosphates from cysteines, which include alkylation of thiophosphates at low pH [4, 10], selective hydrolysis of the P–S thiophosphate bond [11, 12], and immunoaffinity chromatography of alkylated substrates [1]. Here, we describe a protocol that uses thiol-ene chemistry to chemoselectively cap proteinaceous thiols, enabling the specific tagging of thiophosphates [13]. The thiol-ene reaction proceeds by a radical-based mechanism and can chemically distinguish thiols from thiophosphates by differences in the electronics and pK_a values of the two functional groups. Photoinitiation instigates the homolytic cleavage of thiol S–H bonds; the resulting sulfur radicals are successively coupled to an alkene, "capping" the thiols. We selected lithium phenyl-2,4,6-trimethylbenzoylphosphinate (LAP) as the photoinitiator and allyl alcohol ("ene") as the capping reagent for our studies, largely due to the water solubility of these reagents (*see* Fig. 1). After the thiol-ene reaction has blocked cysteine residues, the preserved thiophosphates are tagged with a fluorescent iodoacetamide reagent by nucleophilic displacement. Labeling experiments using our thiol-ene protocol (*see* Fig. 2) successfully visualized thiophosphorylation between JNK1/ATF-2, p38α/MBP, and Src autothiophosphorylation. Substrate detection limits mirrored those of previous studies and were linear with concentration. In addition, we determined that the thiol-ene reaction blocks over 85 % of protein thiols [13]. While this protocol focuses on fluorescence, other iodoacetamide variants (i.e., biotinylated) could be used to enrich or detect thiophosphoproteins [4, 10].

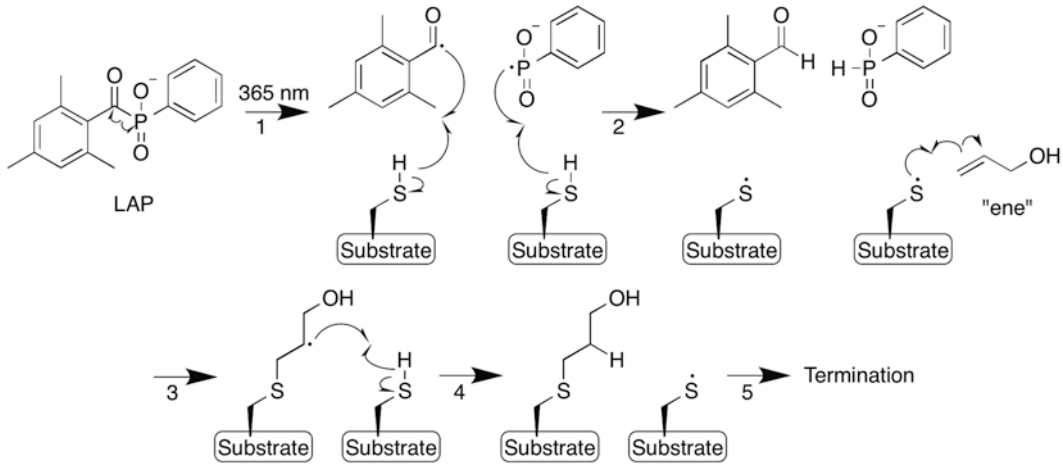


Fig. 1 Thiol-ene reaction mechanism for cysteine capping: Upon irradiation with 365 nm light, the photoinitiator LAP dissociates into two radical species (**1**) that initiate a radical chain reaction through homolysis of thiol S–H bonds (**2**). The generated thiyl radicals react with allyl alcohol (the "ene") (**3**), and the chain reaction propagates to generate new thiyl and ene radicals, for which one example is shown (**4**). The process continues until two free-radical species couple, resulting in termination (**5**) (refs. 17, 22). In this way, cysteine residues are capped

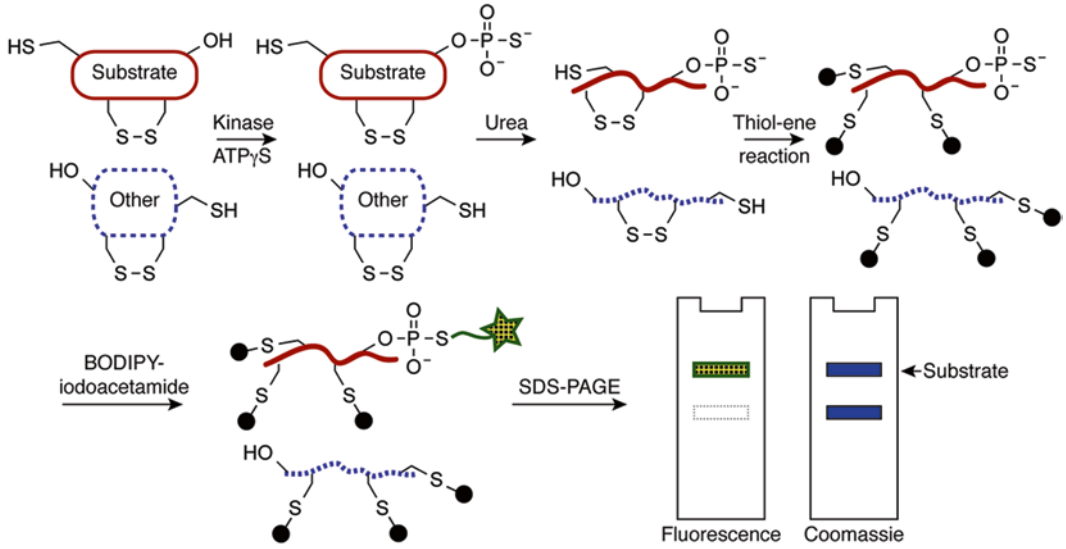


Fig. 2 Thiol-ene-enabled detection of thiophosphorylated substrates: The proteome of interest is first reacted with ATPγS, which kinases utilize to thiophosphorylate their substrates. Urea denatures the proteins to expose cysteine residues. Using the thiol-ene reaction, both reduced and oxidized thiols are chemoselectively capped. Lastly, a fluorescent tag reacts with the thiophosphates by nucleophilic displacement. In-gel fluorescence detection illuminates thiophosphoproteins only, in comparison to Coomassie staining in which all proteins are visualized

2 Materials

Prepare all solutions using Milli-Q purified water. Those described under components for thiophosphorylation (*see* Subheading 2.3) and thiol-ene chemistry (*see* Subheading 2.4) should be prepared fresh daily and stored on ice until use. Follow all disposal regulations upon discarding waste. Volumes provided in the protocol are sufficient for 15 reactions to load a single 15-well protein gel.

2.1 Sodium Dodecyl Sulfate Polyacrylamide Gel

1. Resolving gel buffer: 1.5 M Tris–HCl, pH 8.8.
2. Stacking gel buffer: 0.5 M Tris–HCl, pH 6.8.
3. Ammonium persulfate: 10 % (w/v) solution in water. After preparation, store at 4 °C, and use within 1 week.
4. *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED). Store at 4 °C.
5. Acrylamide/Bis-acrylamide (29:1): 40 % solution (*see* **Note 1**).
6. Empty gel cassettes, mini, 1.5 mm.
7. Empty gel cassette combs, mini, 1.5 mm, 15 well.
8. XCell *SureLock* Mini-Cell electrophoresis chamber.
9. SDS-PAGE running buffer: 10× buffer. Dilute tenfold to achieve 1× buffer.

10. 2× SDS-PAGE sample loading buffer: 125 mM Tris–HCl, pH 6.8, 20 % glycerol (v/v), 4 % SDS (w/v), 5 % BME (v/v), and 0.2 % bromophenol blue (w/v). Store aliquots at −20 °C. Thaw completely prior to use.
11. Benchmark protein ladder.
12. Plastic tray large enough to wash 8 cm × 8 cm gel.
13. Kimwipes.
14. 50 mL Falcon tubes.

2.2 Background Proteome

1. *Escherichia coli* K12 (*see* **Note 2**).
2. Sterile lysogeny broth (LB).
3. Sterile test tubes.
4. Microcentrifuge tubes.
5. Centrifuge.
6. Microcentrifuge.
7. Sonifier.

2.3 Thiophosphorylation

1. Reaction buffer: 50 mM Tris–HCl, pH 7.8, 200 mM KCl, 5 mM $MgCl_2$. Prepare a 5× reaction buffer (250 mM Tris–HCl, pH 7.8, 1 M KCl, 25 mM $MgCl_2$). As needed, dilute 10.0 mL of the 5× buffer with 40.0 mL water. Adjust pH back to 7.8 using hydrochloric acid (HCl) or potassium hydroxide (KOH) if necessary. Sterile filter using a pore size of 0.22 μm.
2. Purified kinase and corresponding substrate can be purchased from various vendors (*see* **Note 3**).
3. ATPγS (tetralithium salt) (Tocris Bioscience): Store reagent bottle at −20 °C. To prepare a 25 mM working stock solution, mix 0.7 mg with 51.2 μL reaction buffer. Confirm the concentration by spectrophotometry (*see* **Note 4**).
4. Nucleotide competitor: ATP or adenosine diphosphate (ADP) (*see* **Note 5**). Prepare a saturating concentration to compete with ATPγS in the final reactions (*see* **Note 6**).
5. Mini microcentrifuge for quick spins, such as the Galaxy Ministar.

2.4 Thiol-ene Reaction

1. 16.7 M urea: Dissolve 801.0 mg urea in 800 μL reaction buffer.
2. 125 mM lithium phenyl-2,4,6-trimethylbenzoylphosphinate (LAP): Synthesize LAP according to Fairbanks et al. (ref. 14) (*see* **Note 7**). Dissolve 1.4 mg in 39.0 μL reaction buffer.
3. 10 M allyl alcohol ("ene"): Mix 34.0 μL allyl alcohol (*see* **Note 8**) with 16.0 μL reaction buffer.

4. 250 mM glutathione: Dissolve 19.2 mg reduced glutathione in 250 μL reaction buffer.
5. UV lamp: Lamp capable of irradiation at 365 nm (*see* **Note 9**).

2.5 Protein Precipitation

1. Trichloroacetic acid (TCA; 100 % w/v): Chill ~100 μL on ice (*see* **Note 10**).
2. Methanol: Chill ~800 μL on ice.
3. Acetone: Chill ~6 mL on ice.

2.6 Thiophosphate Tagging

1. Dimethyl sulfoxide (DMSO).
2. BODIPY-iodoacetamide: Prepare at least 300 μL of 1 μM BODIPY-iodoacetamide by diluting a DMSO freezer stock (*see* **Note 11**) with reaction buffer (*see* **Note 12**).

2.7 Protein Detection

1. Fluorescence gel scanner (*see* **Note 13**).
2. Gel quantitation software (*see* **Note 14**).
3. Coomassie stain: 0.1 % (w/v) Coomassie Brilliant Blue R-250, 10 % acetic acid, 40 % methanol, and 50 % water.
4. Destain: 10 % acetic acid, 40 % methanol, and 50 % water.

3 Methods

Carry out procedures at room temperature unless otherwise specified.

3.1 SDS-PAGE Preparation

1. In a conical tube, prepare a 12.5 % resolving gel by adding 3 mL resolving buffer, 3.75 mL acrylamide solution, 5.25 mL water, 42 μL ammonium persulfate, and 21 μL TEMED (*see* **Note 15**). Invert 15–20 times to mix, avoiding bubbles. Into a cassette or plates/spacers with dimensions of 8 cm × 8 cm × 1.5 mm, pour the gel solution, but leave about 20 % of the upper space for a stacking gel (*see* **Note 16**). Overlay the gel solution with 100 % ethanol. Allow gel to polymerize for 1 h.
2. Once polymerized, pour off ethanol. Rinse the top of the resolving gel with water three times. Use strips of paper towel to dry between the plates and the top of the gel.
3. In a conical tube, prepare the 4.5 % stacking gel by adding 2.5 mL stacking buffer, 1.1 mL acrylamide solution, 6.4 mL water, 30 μL ammonium persulfate, and 10 μL TEMED. Mix gently by inverting 15–20 times. Pour mixture over the top of the resolving gel to fill the remaining space in the cassette. Carefully insert a 15-well comb so as not to introduce air bubbles. Stacking gel will polymerize in 2–3 h (*see* **Note 17**).

3.2 Preparation of Background Proteome

1. Grow *E. coli* K12 cells in 5 mL lysogeny broth (LB) to mid-log phase.
2. Collect cells by centrifugation at 4000 × *g* for 5 min.
3. Resuspend cells in 1 mL reaction buffer to wash them, and transfer to microcentrifuge tube. Centrifuge to pellet cells, and wash once more. Resuspend cells in 300 μL reaction buffer.
4. Lyse cells using a Branson Sonifier 250 with 1/8 in. tapered microtip (power setting 3, 30 % duty cycle, 5 min on ice).
5. Collect the soluble proteome by centrifugation at 21,000 × *g* for 10 min at 4 °C.
6. Transfer supernatant to a new tube and measure the concentration in mg/mL on a NanoDrop spectrophotometer. Adjust the concentration to 5–10 mg/mL with reaction buffer (*see* **Note 18**).

3.3 Kinase-Catalyzed Thiophosphorylation of Substrates

1. For each reaction in a microcentrifuge tube, mix 0.1–3 μg kinase, 5–10 μg substrate, and 1 μL of 25 mM ATPγS (*see* **Note 19**) to a final volume of 5 μL in reaction buffer. If a competitor is to be analyzed, include that in the 5 μL volume (*see* **Note 20**). Mix using a vortex mixer, and collect the reaction solution at the bottom of the tube with a mini microcentrifuge (*see* **Note 21**).
2. Incubate reactions at 37 °C for 12 h (*see* **Note 22**).

3.4 Thiol-ene Reaction for Selective Capping of Thiols

1. Set reactions on bench to bring back to room temperature.
2. Add 20 μg background proteome (*see* **Note 23**) and 18 μL 16.7 M urea (*see* **Note 24**) to a total volume of 30 μL in reaction buffer. Mix well by vortex mixer, centrifuge for 1–2 s, and let incubate at room temperature for 10 min.
3. Add 10 μL 250 mM reduced glutathione (*see* **Note 25**), 2 μL 125 mM LAP (*see* **Note 26**), and 1 μL 10 M allyl alcohol (*see* **Note 27**). Add reaction buffer (in this case, 7 μL) to bring to a final volume of 50 μL (*see* **Note 28**). Mix well by vortex mixer, and centrifuge for 1–2 s.
4. Using the UV lamp (*see* **Note 29**), irradiate samples at 365 nm for 20 min on ice (*see* **Note 30**).

3.5 Protein Precipitation

1. On ice, add 5 μL cold TCA to each sample. Vortex, and incubate on ice for 10 min.
2. Centrifuge samples at 4 °C for 10 min at 20,000 × *g*. Discard the supernatant (*see* **Note 31**).
3. Add 40 μL cold methanol to wash pellets. Mix well by vortex mixer, and centrifuge samples at 4 °C for 10 min at 20,000 × *g*. Discard the supernatant.

4. Add 350 μL acetone to wash away methanol. Mix well by vortex mixer, and centrifuge samples at 4 °C for 10 min at 20,000 × *g*. Discard the supernatant (*see* **Note 32**).
5. Set tubes upside down on a paper towel to dry at room temperature for 30 min.

3.6 Protein Thiophosphate Tagging

1. Add 15 μL of 1 μM BODIPY-iodoacetamide to each sample (*see* **Note 33**). Mix well by vortex mixer, and centrifuge for 1–2 s. Incubate reactions at room temperature for 1 h in the dark to prevent photobleaching.
2. Quench reactions with 5 μL 2× SDS-PAGE sample loading buffer. Mix by vortex mixer, and centrifuge for 1–2 s.
3. Heat samples to 95 °C for 5 min. Cool to room temperature, and mix well again.

3.7 SDS-PAGE

1. Load 16 μL of each sample onto the SDS-PAGE gel (*see* **Note 34**). Add 8 μL Benchmark Protein Ladder to one lane.
2. Run at 180 V, 400 mA, and 60 W for 1 h or until dye front reaches the bottom of the gel.
3. Pry open gel cassette plates, and cut the corner closest to lane 1. Trim off the dye front (*see* **Note 35**), and transfer gel to a tray containing water. Wash the gel by rocking back and forth, and replace the tray with fresh water three times to remove SDS.

3.8 In-Gel Fluorescence Detection

1. Scan the gel using a fluorescence gel imager. If using a Typhoon Variable Mode Imager (*see* **Note 13**), select the 526 nm short-pass filter to detect BODIPY fluorescence (λ_{ex} 504 nm, λ_{em} 514 nm). Set the photomultiplier tube (PMT) to 525 V (*see* **Note 36**) and the scan resolution to 50 μm.
2. Using image-analysis software, integrate the density of the fluorescent gel bands, which represent kinase-catalyzed thiophosphorylated substrates (*see* **Note 37**).
3. After fluorescence imaging, place the gel in a tray with just enough Coomassie stain to cover it, and incubate with agitation for 10 min. Discard stain.
4. Fill the tray with destain, and incubate with agitation for 30 min (*see* **Note 38**). Discard destain.
5. Fill the tray with water, and incubate with agitation overnight. The next day, analyze the material (*see* **Note 39**) visualized by Coomassie to ensure that protein was evenly loaded onto the gel in all lanes (*see* **Note 40**) (*see* Fig. 3).

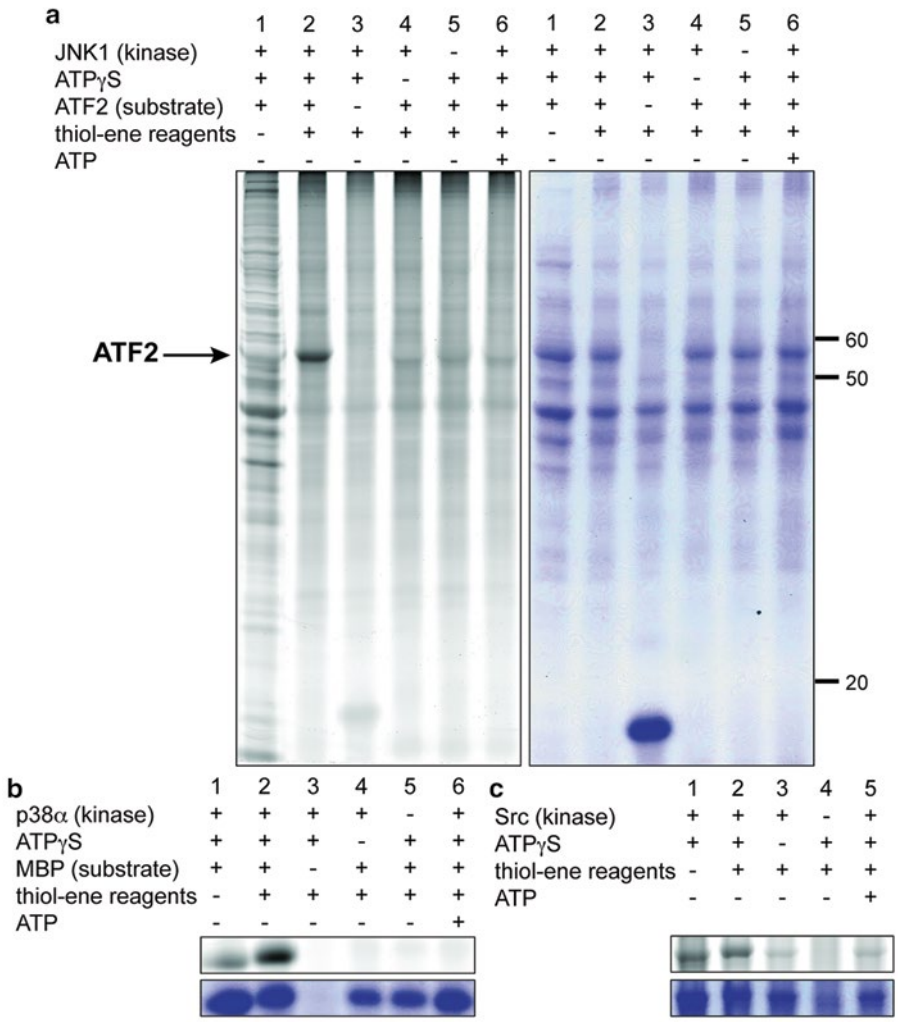


Fig. 3 Fluorescent labeling of thiophosphorylated JNK1 (**a**), p38α MAPK (**b**), and Src (**c**): (**a** and **b**) The serine/threonine kinase-substrate pairs JNK1-ATF2 and p38α-MBP were incubated with ATPγS (80 ng kinase, 8 μg substrate). In-gel fluorescence detection of the BODIPY fluorophore in the absence (*lane 1*) or presence (*lane 2*) of treatment with the thiol-ene reagents. Substrate labeling is abolished in the absence of ATPγS (*lane 4*) or kinase (*lane 5*), as well as when preincubated with ATP (*lane 6*; 100 mM) (*see* **Note 20**). *Lane 3* in *panel* (**a**) contains lysozyme (14 kDa), which was required to facilitate effective protein precipitation in the absence of substrate ATF2. (**c**) The autophosphorylating tyrosine kinase Src was also examined (2.4 μg). Coomassie staining showed even protein loading (*see* **Note 40**). Reprinted with permission from Garber, K. C. A., and Carlson, E. E. (2013) Thiol-ene enabled detection of thiophosphorylated kinase substrates. *ACS Chem. Biol. 8*, 1671–1676. Copyright 2013 American Chemical Society [13]

4 Notes

1. Use caution when handling the acrylamide solution. It is a toxic reagent and suspected carcinogen. Wear personal protective equipment (PPE) when preparing gels.

2. Any cell lysate will be appropriate. We have used both bacterial (*E. coli* K12) and eukaryotic (human breast tumor cell line MCF-7) cell lysates. The preparation of *E. coli* proteome is provided.
3. Examples of kinase-substrate pairs include the following: JNK1 (kinase)/ATF-2 (substrate); p38α (kinase)/MBP (substrate); Src (autokinase). If provided as a lyophilized powder, follow vendor data sheets and certificates of analysis for reconstitution instructions. Avoid repeated freezing and thawing; storing as small aliquots will minimize this.
4. Using the UV–Vis option on a NanoDrop spectrophotometer, measure the absorbance of the ATPγS solution at 259 nm, using reaction buffer as a blank. A millimolar ATPγS stock solution is too concentrated to fall within the linear range of the instrument, and the resulting spectrum will appear jagged due to saturation. To accurately measure, prepare 1:100 and 1:1000 dilutions of the stock in reaction buffer. Measure each at 259 nm. Using the extinction coefficient (15,400 M^{-1} cm^{-1}) and Beer's law (1), solve for the concentration of the diluted ATPγS solutions and original working stock. Beer's law is defined as

$$A = \varepsilon c \ell \quad (1)$$

where A is the absorbance, ε is the extinction coefficient, c is the molar concentration, and l is the pathlength in centimeter. Note that the NanoDrop pathlength will be 1.0 mm (not 1.0 cm). If the solution is not 25 mM, adjust reagents or later volumes to accommodate the actual concentration. This will be used to prepare a final concentration of 5 mM in the reactions.

5. Various nucleotide analogues that are commercially available could be used as competitors. We commonly use ATP and ADP. Likewise, if specific inhibitors exist for a kinase-substrate pair, those could also be used.
6. We have found an ATP concentration 20-fold that of ATPγS to be sufficient. Confirm ATP concentration as done with ATPγS (*see* **Note 3**) ($\varepsilon = 15,400\ M^{-1}\ cm^{-1}$).
7. Store LAP at −80 °C under dry conditions. Transferring small amounts into tubes and noting the mass of LAP in each can make solution preparation faster later.
8. Allyl alcohol is a liquid with a density of 0.854 g/mL. This reagent comes with a *SureSeal* to prevent contamination by oxygen and water. To remove reagent from bottle, first flush an inlet line and needle with argon for 1–2 min. Insert into septum at a low argon flow rate. Insert a syringe into the bottle, and remove 5–10 mL allyl alcohol. Aliquot into tubes, and store in a container with desiccant at −80 °C until needed.

Note that allyl alcohol is hazardous (flammable and toxic). Follow appropriate precautions, and wear PPE.

9. We use a UVP Black-Ray UV Bench Lamp XX-15L, 365 nm, 115 V, 60 Hz, 0.68 Amps, 15 W.
10. TCA is hazardous and corrosive, so proper PPE should be adhered when handling acid.
11. Prepare the BODIPY-iodoacetamide stock by reconstituting the vial's entire 5 mg contents in 480 μL DMSO to make a 25 mM stock. Keep dry, protect from light, and store at −80 °C. Preparing aliquots can prevent repeated freeze-thaw cycles.
12. DMSO may adhere to pipette tips. When pipetting up DMSO solutions, place tip just on the surface of the liquid, and avoid submerging tip largely into the solution.
13. We use the Typhoon Variable Mode Imager 9210.
14. We use ImageJ (NIH). This can be downloaded from the following website: http://imagej.nih.gov/ij/ (ref. 15).
15. This recipe is for a 12.5 % resolving gel, which should adequately separate proteins with molecular weights ranging from 14 to 200 kDa. However, be sure to adjust the gel percentage if proteins of other sizes are to be analyzed or if resolution at high molecular weights is not sufficient.
16. Gel recipe scales up well if several gels are needed.
17. If multiple gels are prepared, wrap them in wet paper towels. Store in zip-top bags with extra water at 4 °C.
18. To prepare background proteome for multiple experiments, larger cultures can be grown and cell pellet stored as aliquots at −80 °C until needed.
19. A final concentration of 5 mM ATPγS is desired.
20. A sample utilizing a competitor (*see* **Note 5**) is an appropriate negative control. At the end of the experiment, lack of a fluorescent band for this sample will ensure that detection of the thiophosphorylation labeling was dependent on activity rather than a result of nonspecific labeling.
21. Because the reactions employ small volumes, it is important to collect the liquid that scatters on the inside of the tubes after vortexing. Use a small microcentrifuge to spin for 1–2 s.
22. In our original studies we incubated proteins with ATPγS for approximately 12 h, which we suggest to start. Later experiments in our laboratory with other proteins have had successful labeling in as short as 1 h, so incubation may be optimized depending on the goal of the experiment and for different proteins.

23. The added protein concentration from the background proteome facilitates protein precipitation (*see* Subheading 3.5) and deters nonspecific fluorophore labeling of kinase and substrate (*see* Subheading 3.6).
24. A concentration of 6 M urea is desired at a final volume of 50 μL after the remaining thiol-ene reagents have been added (*see* Subheading 3.4, **step 3**). The urea-induced denaturation helps expose buried thiols. The addition of reducing agent is not necessary as thiol-ene chemistry effectively modifies both reduced and oxidized (i.e., disulfide) thiols (refs. 16, 17).
25. A concentration of 50 mM glutathione is desired in a final 50 μL volume. The procedure was optimized to include glutathione because it was found to help minimize side reactions and cross-linking due to the thiol-ene chemistry.
26. A concentration of 5 mM LAP is desired in a final 50 μL volume. While many photoinitiators exist, LAP was selected because it is soluble in aqueous solutions (refs. 13, 14). Additionally, LAP had been successful in other protein-based thiol-ene experiments (refs. 18–20).
27. Allyl alcohol was chosen as the "ene" for its water solubility. A final concentration of 200 mM is desired in 50 μL.
28. Reagents must be added individually if performing controls (e.g., excluding single reagents). However, if each sample is to contain all reagents, it may be easier to prepare a "master mix" containing background proteome, urea, glutathione, LAP, allyl alcohol, and reaction buffer (*see* Subheading 3.4, **steps 2** and **3**). Therefore, only one pipetted volume from this "master mix" needs to be added per sample.
29. Situate microcentrifuge tubes on their sides and close to the bulb of the lamp (within 2–3 cm). It is important to keep the tubes of equal distance from the light source to ensure that light-mediated activation is consistent among samples. If using a lamp of lower wattage, additional irradiation time may be necessary.
30. We have found that adding the "ene" in two separate steps, followed by irradiation each time, yields more successful thiol capping. In this case, add 0.5 μL allyl alcohol, mix, and irradiate for 10 min. Subsequently add the remaining 0.5 μL allyl alcohol, mix, and irradiate for another 10 min.
31. The acid will precipitate proteins, which collect as a pellet during centrifugation. Unreacted ATPγS, hydrolyzed thiophosphate, and glutathione will remain in the supernatant for removal. Supernatant can be poured out of the microcentrifuge tube or can be carefully pipetted off by placing the tip on the side of the tube opposite from the pellet.

32. The pellet will be visible after TCA precipitation. After the acetone wash, however, it will be difficult to see. To remove all acetone, tubes can be turned over and tapped on a paper towel.
33. We also tried thiophosphate tagging with BODIPY-maleimide; however, the iodoacetamide reagent yielded cleaner gel bands.
34. This corresponds to approximately 0.08–2.4 μg kinase and 4–8 μg substrate per lane. It may help to pipette up and down prior to transferring 16 μL from the tube to the well in case proteins have settled on the bottom of the tube.
35. Excess fluorophore will migrate close to the dye front. If left untrimmed, this will add a lot of fluorescent background to the bottom of the gel during imaging.
36. The PMT voltage can be increased to improve signal collection.
37. Background should be subtracted to discount background fluorescence. Additionally, the brightness and contrast can be adjusted to visually improve the signal-to-noise ratio. However, any changes must be made to the entire gel image such that each lane is comparable. If there is heavy fluorescence smearing in the gel lanes, one of the following may be helpful: decrease the concentration of BODIPY-iodoacetamide used to label thiophosphorylated proteins, load less sample on the gel, or decrease the PMT voltage on the gel imager.
38. Putting a few Kimwipes in the tray will help soak up Coomassie from the gel during the destaining process.
39. The Benchmark Protein Ladder will only be visible by Coomassie staining. Other protocols have established the preparation of fluorescent molecular weight markers if that is desired (ref. 21).
40. The fluorescent bands indicate activity due to the detection of thiophosphorylation. However, fluorescence intensity can also be affected by the amount of protein loaded. Coomassie staining is important as it confirms that differences in fluorescence intensity between samples can be attributed to changes in activity and not protein level.

Acknowledgements

We thank Kathleen C. A. Garber for helpful discussions. This work was supported by NIH DP2OD008592, a Pew Biomedical Scholar Award (E.E.C.), the Research Corporation for Science Advancement (Cottrell Scholar Award) (E.E.C.), a Sloan Research Fellowship (E.E.C), and an Indiana University Quantitative and Chemical Biology Training Fellowship (K.E.W.).

References

1. Allen JJ, Lazerwith SE, Shokat KM (2005) Bio-orthogonal affinity purification of direct kinase substrates. J Am Chem Soc 127:5288–5289
2. Allen JJ, Li M, Brinkworth CS et al (2007) A semisynthetic epitope for kinase substrates. Nat Methods 4:511–516
3. Kinoshita E, Kinoshita-Kikuta E, Shiba A et al (2014) Profiling of protein thiophosphorylation by Phos-tag affinity electrophoresis: evaluation of adenosine 5'-O-(3-thiotriphosphate) as a phosphoryl donor in protein kinase reactions. Proteomics 14:668–679
4. Kwon SW, Kim SC, Jaunbergs J et al (2003) Selective enrichment of thiophosphorylated polypeptides as a tool for the analysis of protein phosphorylation. Mol Cell Proteomics 2:242–247
5. Carlson SM, White FM (2012) Labeling and identification of direct kinase substrates. Sci Signal 5:pl3
6. Lourido S, Jeschke GR, Turk BE et al (2013) Exploiting the unique ATP-binding pocket of Toxoplasma Calcium-Dependent Protein Kinase 1 to identify its substrates. ACS Chem Biol 8:1155–1162
7. Uhalte EC, Kirchner M, Hellwig N et al (2012) In vivo conditions to identify Prkci phosphorylation targets using the analog-sensitive kinase method in zebrafish. PLoS One 7:e40000
8. Cassel D, Glaser L (1982) Resistance to phosphatase of thiophosphorylated epidermal growth factor receptor in A431 membranes. Proc Natl Acad Sci U S A 79:2231–2235
9. Hiriyanna KT, Baedke D, Baek KH et al (1994) Thiophosphorylated substrate analogs are potent active site-directed inhibitors of protein-tyrosine phosphatases. Anal Biochem 223:51–58
10. Lee SE, Elphick LM, Kramer HB et al (2011) The chemoselective one-step alkylation and isolation of thiophosphorylated Cdk2 substrates in the presence of native cysteine. ChemBioChem 12:633–640
11. Blethrow JD, Glavy JS, Morgan DO et al (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. Proc Natl Acad Sci U S A 105:1442–1447
12. Chi Y, Clurman BE (2010) Mass spectrometry-based identification of protein kinase substrates utilizing engineered kinases and thiophosphate labeling. Curr Protoc Chem Biol 2:219–234
13. Garber KC, Carlson EE (2013) Thiol-ene enabled detection of thiophosphorylated kinase substrates. ACS Chem Biol 8:1671–1676
14. Fairbanks BD, Schwartz MP, Bowman CN et al (2009) Photoinitiated polymerization of PEG-diacrylate with lithium phenyl-2,4,6-trimethylbenzoylphosphinate: polymerization rate and cytocompatibility. Biomaterials 30:6702–6707
15. Rasband WS (1997–2014) ImageJ. U.S. National Institutes of Health, Bethesda, MD
16. Lo Conte M, Staderini S, Marra A et al (2011) Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. Chem Commun 47:11086–11088
17. Fairbanks BD, Singh SP, Bowman CN et al (2011) Photodegradable, photoadaptable hydrogels via radical-mediated disulfide fragmentation reaction. Macromolecules 44:2444–2450
18. Dondoni A, Marra A (2012) Recent applications of thiol-ene coupling as a click process for glycoconjugation. Chem Soc Rev 41:573–586
19. Li YM, Yang MY, Huang YC et al (2012) Genetically encoded alkenyl-pyrrolysine analogues for thiol-ene reaction mediated site-specific protein labeling. Chem Sci 3:2766–2770
20. Valkevich EM, Guenette RG, Sanchez NA et al (2012) Forging isopeptide bonds using thiol-ene chemistry: site-specific coupling of ubiquitin molecules for studying the activity of isopeptidases. J Am Chem Soc 134:6916–6919
21. Kocaoglu O, Carlson EE (2013) Penicillin-binding protein imaging probes. Curr Protoc Chem Biol 5:239–250
22. Griesbaum K (1970) Problems and possibilities of the free-radical addition of thiols to unsaturated compounds. Angew Chem Int Ed Engl 9:273–287

Chapter 2

Phosphopeptide Detection with Biotin-Labeled Phos-tag

Emiko Kinoshita-Kikuta, Eiji Kinoshita, and Tohru Koike

Abstract

Protein kinases are widely considered to be invaluable target enzymes for drug discovery and for diagnosing diseases and assessing their prognosis. Effective analytical techniques for measuring the activities of cellular protein kinases are therefore required for studies in the field of phosphoproteomics. We have recently developed a highly sensitive microarray-based technique for tracing the activities of protein kinases. A series of peptides that are specific substrates of various protein kinases are immobilized on a glass slide and subjected to phosphorylation by cell lysates. The resulting phosphorylated forms of the various peptides are then selectively and simultaneously detected by using a phosphate-binding tag molecule, biotin-labeled Phos-tag, bound to horseradish peroxidase-conjugated streptavidin. Enhanced chemiluminescence signals can then be readily detected by using an automatic image analyzer. In this chapter, we describe a standard protocol for detecting phosphopeptides by biotin-labeled Phos-tag. We also describe a microarray system for high-throughput profiling of intracellular protein kinase activities. The Phos-tag-based method is expected to be useful in the rapid detection of the complex range of phosphorylation reactions involved in cellular signaling events, and it has potential applications in high-throughput screening of kinase activators or inhibitors.

Key words Phos-tag, Biotin, Protein kinase, Peptide, Microarray, Phosphorylation, High-throughput profiling, Enhanced chemiluminescence

1 Introduction

Protein kinases are ubiquitous catalytic enzymes that rapidly modify their substrate proteins by transferring negatively charged phosphate groups from ATP to serine, threonine, tyrosine, or, less commonly, other residues in the proteins [1]. The phosphorylation reactions cause functional alterations in their targets by changing their enzymatic activity, localization, or specificity of binding to other proteins [2, 3]. More than 500 protein kinase genes are known to be present in the human genome [4]. According to PhosphoSitePlus (http://www.phosphosite.org/homeAction.do), a Web site produced by Cell Signaling Technology, Inc. (Danvers, MA, USA), there are more than 100,000 phosphorylation sites present in human proteins. As a result, up to 80 % of all

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_2,

human proteins (the proteome) can be modified by various activities of protein kinases (the kinome). This posttranslational modification dramatically enhances the diversity of genetically encoded proteins and it provides precise control of many important cellular processes, including signal transduction, gene expression, cell cycle progression, cytoskeletal regulation, and energy metabolism. Dysregulation in kinase activities has been implicated as a key factor in the etiology of many human diseases, such as cancers [5], immune inflammation [6], and neurodegenerative disorders [7].

Effective methods for tracing kinase activities are therefore very important in achieving an understanding of the molecular origins of various diseases and, potentially, in developing tools for therapeutic intervention [8]. A widely used method for the detection of kinase activities is the incorporation of a radioisotope label, such as ^{32}P or ^{33}P, into the phosphorylated targets. The phosphorylation status of the target can then be identified and quantified by measuring its radioactivity. Non-radioisotope-based methods for profiling of phosphorylation reactions by using polyclonal and monoclonal antibodies against phosphorylated amino acid residues are also well established. These methods can be applied in conjunction with many of the standard biochemical analytical techniques, such as filter-binding assays, gel-based autoradiography and immunoblotting assays, or enzyme-linked immunosorbent assays. However, these methods have the disadvantages of providing low throughputs and of requiring complex procedures.

The large-scale identification of protein phosphorylation has recently become possible as a result of dramatic advances in mass spectrometry (MS)-based methods of shotgun proteomics, coupled with improvements in MS instrumentation and the development of better procedures for the enrichment of phosphopeptides [9]. MS-based techniques also provide several strategies, such as isotope-coded affinity tag (ICAT), isobaric tags for relative and absolute quantitation (iTRAQ), or stable isotope labeling by amino acids in cell culture (SILAC), for the quantitative analysis of phosphorylation, particularly the quantitative determination of levels of phosphorylation of individual residues in peptide digests. However, these techniques require expensive instrumentation and involve complicated procedures for sample preparation through enrichment of phosphopeptides after enzymatic digestion.

With regard to the analysis of activities of cellular protein kinases, approaches based on a so-called kinomics microarray, consisting of a series of peptides that are specific substrates of various protein kinases, provide high-throughput, cost-effective, and convenient solutions for taking panoramic snapshots of signal transduction by means of phosphorylation reactions [10, 11]. A number of such peptide microarrays, including PepChip (Pepscan Therapeutics, Lelystad, Netherlands), PepStar (JPT Peptide Technologies GmbH, Berlin, Germany), PamChip (PamGene,

Cambridge, MA, USA), and CelluSpots (Intavis Bioanalytical Instruments AG, Cologne, Germany), are commercially available, and these have been successfully used in studies on the kinomics of cultured cell lines and clinical tissue samples. Levels of phosphorylation of the various substrates immobilized on the microarrays have generally been determined by means of conventional radioisotope- or antibody-based probing methods. Although microarray-based kinomics techniques are inferior to conventional MS-based kinomics techniques in terms of quantitative evaluation, the array technique has unique advantages in terms of its simplicity and its high throughput. It is therefore especially suited to practical analyses of the effects of drugs on the kinome [12].

We have developed a technology known as Phos-tag to permit the analysis of phosphorylated biomolecules. Our Phos-tag technology utilizes a novel phosphate-binding tag molecule, Phos-tag that is capable of capturing phosphate monoester dianions ($R\text{-}OPO_3^{2-}$) in aqueous solutions of neutral pH [13]. The Phos-tag technology has made contributions to the development of several procedures for research on the phosphoproteome, including an immobilized metal (zinc)-affinity chromatography technique for the separation and enrichment of phosphopeptides and phosphoproteins [14–18] and a phosphate-affinity electrophoresis technique for the detection of changes in the mobilities of phosphoproteins in comparison with those of their nonphosphorylated counterparts [19–29]. These techniques use various derivatives of the original Phos-tag molecule. Among these derivatives, biotin-labeled Phos-tag {Phos-tag Biotin; Phos-tag = 1,3-bis[bis(pyridin-2-ylmethyl)amino]propan-2-olato dizinc(II) complex} has been developed as a novel phosphate-affinity probe that has various applications in determining the phosphorylation status of a wide range of peptides and proteins [29–31] (*see* Fig. 1a). Compared with conventional radioisotope- or antibody-based probing methods, our established applications using Phos-tag Biotin offer the following basic advantages.

1. The use of radioactive materials is unnecessary.
2. Because the binding specificity of the Phos-tag molecule is independent of the type of amino acid residue, comprehensive detection of phosphorylation reactions is possible.
3. The application in Western blotting can be succeeded by other downstream procedures, such as antibody reprobing, MS analysis, or Edman sequencing.

We recently demonstrated some useful improvements in techniques for the detection of phosphopeptides and phosphoproteins through the use of a newly synthesized biotin-labeled derivative of Phos-tag that contains an oligomer of 12 molecules of ethylene glycol [PEG12; dodeca(ethylene glycol)] as a long hydrophilic spacer between a biotin moiety and the Phos-tag

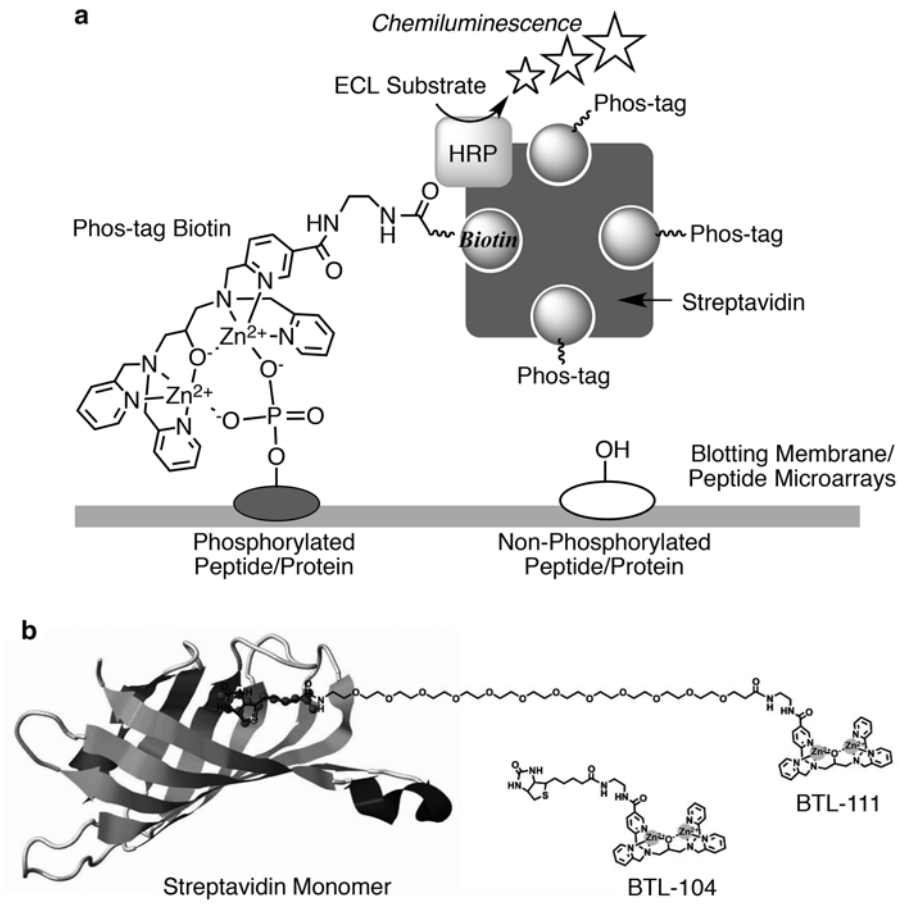


Fig. 1 (**a**) Schematic representation of enhanced chemiluminescence (ECL) detection of phosphorylated peptides or proteins on a peptide-microarray format or a protein-blotting membrane, respectively, by using Phos-tag Biotin. The phosphorylated targets were probed by using the complex of Phos-tag Biotin with HRP-conjugated streptavidin, and then the Phos-tag-bound targets were detected by an ECL system. (**b**) Superimposed images of the complex of a biotin moiety and a streptavidin monomer, and the newly synthesized derivative BTL-111. For reference, the structure of the conventional derivative BTL-104 is also shown. The introduction of the long hydrophilic spacer [PEG12; dodeca(ethylene glycol)] increases the flexibility of the phosphate-affinity moiety of Phos-tag, resulting in a greater sensitivity in the detection of phosphopeptides or phosphoproteins

moiety (Phos-tag Biotin BTL-111) (*see* Fig. 1b) [32]. Because the long spacer confers greater flexibility to the phosphate-affinity Phos-tag moiety, this derivate can access phosphorylated targets more readily, thereby permitting a wider range of applications in specific detections of protein phosphorylation, including microarray-based techniques [33–35]. The advanced Phos-tag-based techniques are therefore expected to permit more-sensitive screening to provide information that might be capable of resolving the nature of complex cellular signaling networks, leading to improved drug discovery and diagnoses of diseases and assessments of their prognoses. Here, we introduce a standard protocol

for detection of phosphopeptides by using Phos-tag Biotin BTL-111 and CelluSpots, a commercially available microarray system, for high-throughput profiling of the kinome activities.

2 Materials (*See* Note 1)

2.1 Preparation of Lysates of Human Cell Lines

1. Culture medium for DLD-1 cells: Dulbecco's modified Eagle medium (DMEM), 10 % (v/v) fetal bovine serum, 100 U/mL penicillin, and 100 μg/mL streptomycin. Store at 4 °C.
2. Culture medium for SW480 cells and A431 cells: RPMI1640 medium, 10 % (v/v) FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin. Store at 4 °C.
3. 1.0 mg/mL epidermal growth factor (EGF) solution for stimulation of A431 cells. Store at −20 °C.
4. Washing Buffer A (Tris-buffered saline; TBS): 10 mM Tris–HCl (pH 7.5) (*see* **Note 2**) and 0.1 M NaCl. Store at room temperature.
5. Radioimmunoprecipitation Assay (RIPA) Buffer: 50 mM Tris–HCl (pH 7.4), 0.15 M NaCl, 0.25 % (w/v) sodium deoxycholate, 1.0 % (v/v) Nonidet P-40, 1.0 mM EDTA. Store at room temperature.
6. Lysis Buffer A: RIPA buffer, 1.0 mM phenylmethanesulfonyl fluoride, 1 μg/mL aprotinin, 1 μg/mL leupeptin, 1 μg/mL pepstatin A, 1 μg/mL leupeptin, 1.0 mM Na_3VO_4, and 1.0 mM NaF. The lysis buffer should be prepared freshly as required and stored on ice before use.
7. Benchtop centrifuge.
8. 1.5-mL microtubes.
9. Protein concentration measurement reagent (e.g., Bio-Rad protein assay).

2.2 Lysate Preparation of Escherichia coli JM109 Strain

1. Culture medium for *Escherichia coli* JM109 cells: Luria–Bertani (LB) broth sterilized by autoclaving. Store at 4 °C.
2. Washing Buffer B: 50 mM NaH_2PO_4–NaOH (pH 8.0) (*see* **Note 3**), 0.3 M NaCl. Store at room temperature.
3. Lysis Buffer B: Washing Buffer B, 10 mg/mL lysozyme. The lysis buffer should be freshly prepared as required and stored on ice before use.
4. 13-mL plastic centrifuge tube.
5. 1.5-mL microtubes.
6. Benchtop centrifuge.
7. Probe sonicator.
8. Protein concentration measurement reagent (e.g., Bio-Rad protein assay).

2.3 On-Chip Kinome Reactions

1. Peptide microarray: CelluSpots (*see* **Note 4**). Store CelluSpots arrays in a cool, dry, dark place at 4 °C (*see* **Note 5**).
2. TBS-T buffer: 10 mM of Tris–HCl (pH 7.5), 0.1 M NaCl, 0.1 % (v/v) poly(oxyethylene) sorbitan monolaurate(Tween). Store at room temperature.
3. Blocking Buffer: TBS-T, 10 % (w/v) bovine serum albumin. The solution should be freshly prepared as required.
4. Washing Buffer C: TBS-T, 1.0 M NaOAc and 1.0 % (w/v) BSA. The solution should be freshly prepared as required.
5. Kinome Reaction Buffer for Tyr kinase substrate arrays: 60 mM HEPES–NaOH, pH 7.5, 10 mM $MgCl_2$, 0.3 mM Na_3VO_4, 2.5 mM dithiothreitol (DTT), 10 mM β-glycerophosphate, 0.2 mM ATP, 1.0 % (w/v) BSA, and 300 μg/mL human cell lysate (or 520 μg/mL bacterial cell lysate) (*see* **Note 6**). If required, add 0.1 μM Src kinase inhibitor I [6,7-dimethoxy-*N*-(4-phenoxyphenyl)quinazolin-4-amine] to the reaction buffer containing the lysate of EGF-stimulated A431 cells. All reaction buffers should be prepared freshly as required and stored on ice before use.
6. Kinome Reaction Buffer for Ser/Thr kinase substrate arrays: 25 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 0.1 mM Na_3VO_4, 2.0 mM DTT, 5.0 mM β-glycerophosphate, 0.2 mM ATP, 1.0 % (w/v) BSA, and 300 μg/mL human cell lysate (or 520 μg/mL bacterial cell lysate). All reaction buffers should be prepared freshly as required and stored on ice before use.
7. 90-mm culture dishes.
8. Orbital shaker.
9. Transparent polypropylene.

2.4 Phosphate-Affinity Probing with Phos-tag

1. Solution A (TBS-T buffer): (*see* Subheading 2.3.2).
2. Solution B (Phos-tag Biotin solution): 1.0 mM Phos-tag Biotin BTL-111 solution is commercially available from Wako Pure Chemical Industries Ltd.. Store in a dark place at 4 °C. The Phos-tag Biotin solution is stable for at least 6 months under these conditions.
3. Solution C [Zinc(II) chloride solution]: 1.0 mM solution of $ZnCl_2$ (e.g., Nacalai Tesque) (*see* **Note 7**). Store at room temperature.
4. Solution D [horseradish peroxidase (HRP)-conjugated streptavidin solution]: HRP–streptavidin conjugate solution from GE Healthcare Bio-Sciences. Store in a dark place at 4 °C.
5. Solution E (4:1 complex of Phos-tag Biotin and HRP-conjugated streptavidin solution): Mix 469 μL of Solution A, 10 μL of Solution B, 20 μL of Solution C, and 1 μL of Solution

D in a 1.5-mL microtube (final volume 500 μL) (*see* **Note 8**) and then allow to stand for 30 min at room temperature.

6. Solution F (phosphate-affinity probing solution): Transfer Solution E (500 μL) to a centrifugal filter device cup (e.g., Spin-X UF 500 Concentrator, 50 K MWCO PES, Corning) and centrifuge at 14,000 × *g* for 20 min at room temperature to remove the excess Zn(II)–Phos-tag Biotin complex (*see* **Note 9**). Dilute the remaining solution (<10 μL) in the cup with 15 mL of a Washing Buffer C solution (*see* Subheading 2.3.4). Solution F should be freshly prepared as required.
7. Plastic bag.
8. Transparent polypropylene.

2.5 Enhanced Chemiluminescence Detection Reagent and Equipment

1. Enhanced chemiluminescence (ECL) detection reagent: Lumigen ECL Ultra. Store in a dark place at 4 °C.
2. Image analyzer: LAS 3000 image analyzer (Fujifilm, Tokyo, Japan).

3 Methods

3.1 Preparation of Lysates from Human Cell Lines

1. To prepare a lysate from DLD-1 cells, incubate the cells (10^7 cells) on a 90-mm culture dish under a humidified atmosphere of 5 % CO_2 and 95 % air at 37 °C overnight (16–24 h), i.e., until all cells have attached to the culture dish.
2. To prepare a lysate from SW480 cells and A431 cells, incubate each cell line (10^7 cells) separately in a 90-mm culture dish with RPMI1640 medium under a humidified atmosphere of 5 % CO_2 and 95 % air at 37 °C overnight (16–24 h), i.e., until all cells have attached to the culture dish.
3. Treat the A431 cells with 250 ng/mL (final concentration) of EGF for 0 min (no treatment) or 5 min.
4. Gently wash the cells attached to the dish twice with 5 mL of Washing Buffer A.
5. Lyse the cells by adding 0.5 mL of Lysis Buffer A to each culture dish.
6. Scrape the lysed cells off and transfer the lysate sample solution to a 1.5-mL microtube.
7. Measure protein concentration using Bio-Rad protein assay.
8. Centrifuge at 14,000 × *g* for 10 min at 4 °C.
9. Transfer the supernatant to a new 1.5-mL microtube and dilute it with an appropriate amount of fresh Lysis Buffer A to give a lysate sample with a protein concentration of 2.0 mg/mL.

10. Mix well and immediately divide into several 50 μL aliquots in 1.5-mL microtubes kept on ice.
11. Store the aliquots (each 50 μL) of the resulting solution at −80 °C until required.

3.2 Preparation of a Lysate from Bacterial JM109 Cells

1. To prepare a lysate from bacterial JM109 cells, incubate 4 μL of competent cells in a 13-mL plastic centrifuge tube with 4 mL of LB Broth medium at 37 °C overnight (16 h).
2. Centrifuge at 1500 × *g* for 3 min at room temperature.
3. Wash the collected cells by suspending them in 0.5 mL of Washing Buffer B and centrifuging at 1500 × *g* for 3 min at room temperature.
4. Suspend the collected cells in 0.5 mL of Lysis Buffer B and allow the mixture to stand on ice for 30 min.
5. Sonicate the mixture on ice until it becomes clear and centrifuge at 14,000 × *g* for 10 min at 4 °C.
6. Measure protein concentration using Bio-Rad protein assay.
7. Transfer the supernatant to a 1.5-mL microtube and dilute it with an appropriate amount of fresh Lysis Buffer B to give a lysate sample with a protein concentration of 2.5 mg/mL.
8. Mix well and immediately divide into several 50 μL aliquots in 1.5-mL microtubes kept on ice.
9. Store the 50 μL aliquots of the resulting solution at −80 °C until required.

3.3 On-Chip Kinome Reactions

1. To prevent nonspecific binding, block the peptide microarrays by immersing each glass slide in 10 mL of Blocking Buffer in a 90-mm culture dish for 4 h at room temperature (or overnight at 4 °C) with gentle shaking on an orbital shaker (*see* **Note 10** and Fig. 2a, left-hand panel).
2. Wash the arrays twice by immersing each glass slide in 10 mL of Washing Buffer C in a 90-mm culture dish for 5 min at room temperature on an orbital shaker.
3. With a pipette, drop 200 μL of the Kinome Reaction Buffer containing the human cell lysate onto the arrays on the slide (*see* **Note 11**); in the case of the bacterial lysate, 240 μL is used.
4. Cover the glass slide with a thin, flat, and rectangular (~120 × 100 mm) piece of transparent polypropylene sheet to prevent the reaction solution from drying, and then incubate the slide at room temperature for 1 h.
5. After incubation, carefully remove the cover sheet.
6. Wash the arrays twice by immersing each glass slide in 10 mL of Washing Buffer C in a 90-mm culture dish for 5 min at room temperature on an orbital shaker.

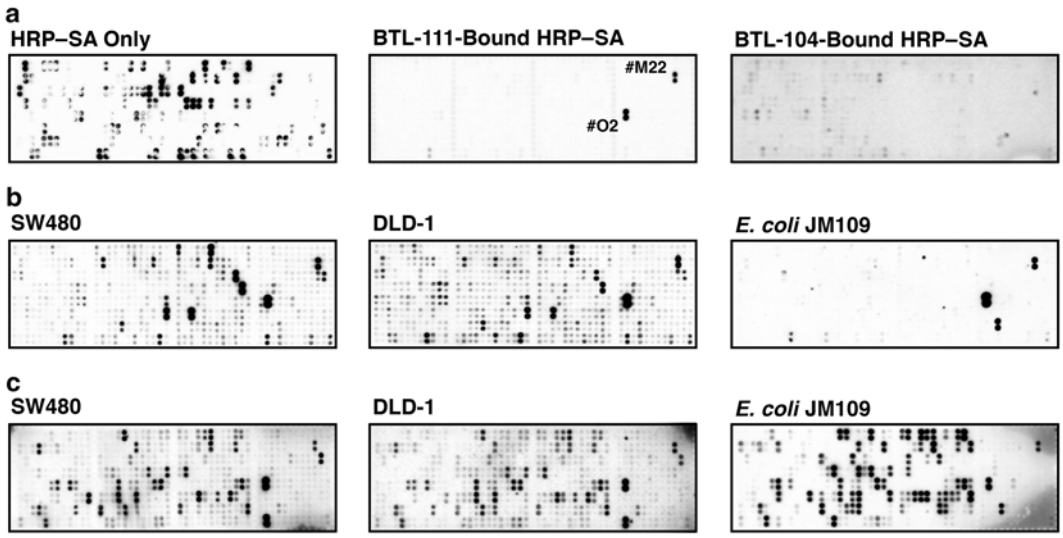


Fig. 2 High-throughput profiling of kinome activities by using a peptide microarray system. (**a**) Comparative data obtained by probing with HRP-conjugated streptavidin (HRP–SA) only (*left*), with the complex of BTL-111 and HRP–SA (*center*), and with the complex of BTL-104 and HRP–SA (*right*). The locations of control spots of phosphopeptides (spots #M22 and #02) are shown in the *center panel*. We first probed the peptide microarray before the kinome reaction by using HRP–SA only. This resulted in the detection of many false-positive ECL signals (*left-hand panel*). To overcome this problem, we incorporated a blocking treatment. We then compared the potency of BTL-111 and BTL-104 with respect to the specific detection of control phosphopeptides after the blocking treatment. The comparative results showed that BTL-111 permitted the specific detection of the control phosphopeptides (*center panel*), whereas the conventional BTL-104 derivative did not do so (*right-hand panel*). The result for BTL-104 indicated that the presence of the blocking protein on the array surface interfered with access to the target by BTL-104. We concluded that the presence of a long spacer in BTL-111 is very important for capturing phosphorylated targets without steric hindrance in this microarray application. Profiling of kinome activities by using the Tyr kinase substrate arrays (YKS-I) (**b**) and Ser/Thr kinase substrate arrays (STKS-II) (**c**) for cell lysates of human colorectal adenocarcinoma SW480 (*left*), human DLD-1 (*center*), and bacterial *Escherichia coli* (JM109 strain). After the blocking treatment with BSA, the samples of the two human cell lysates (60 μg proteins each) and the bacterial lysate (125 μg proteins) were independently applied to the peptide microarray. For both biological species, BTL-111 was suitable for performing high-throughput assays with highly selective and sensitive detection of kinome activities. The bacterial sample showed fewer Tyr kinase activities than did the human samples. This reflects the expression level of the eukaryotic-like Tyr kinase family in the bacterial cell. Reproduced from ref. 32 with permission from WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

3.4 Phosphate-Affinity Probing with Phos-tag and for Detecting Phosphopeptides

1. After washing, independently incubate each glass slide with 5 mL of Solution F in a plastic bag. Gently rock the bag on an orbital shaker for 30 min at room temperature.
2. Carefully remove the glass slide from the bag and wash it twice by immersing it in 10 mL of Washing Buffer C in a 90-mm culture dish for 5 min at room temperature on an orbital shaker.
3. For ECL detection, treat each glass slide with an appropriate amount of the ECL reagent according to the manufacturer's instructions.
4. Detect the ECL signals by using the LAS 3000 image analyzer (*see* Figs. 2 and 3) (*see* **Note 12**).

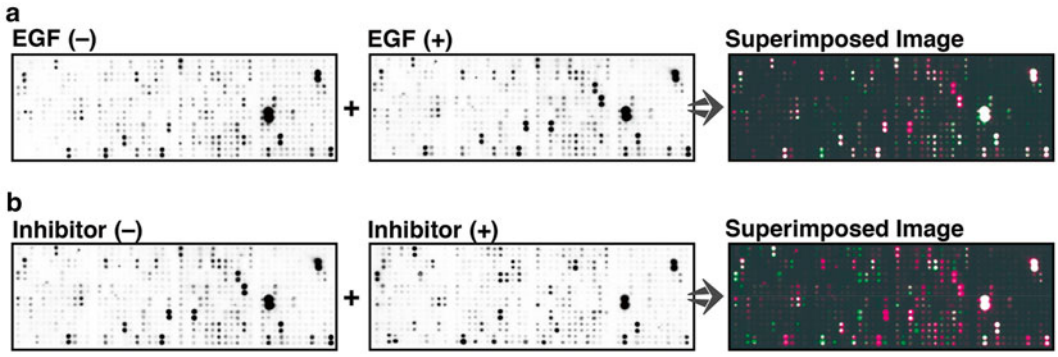


Fig. 3 High-throughput profiling of intracellular Tyr kinase activities of EGF-stimulated A431 cells by using YKS-I arrays. (**a**) The images of detections from lysates before (control, –) and after (+) EGF stimulation are shown in the *left-hand* and *center panels*, respectively. The two images are superimposed in the *right-hand panel*. The ECL signals from the control spots #M22 and #O2 were used to normalize the two different ECL images. The image obtained by detection using the lysate before EGF stimulation is represented by *green color* and the image obtained by detection using the lysate after EGF stimulation is represented by *magenta color*. When these two images were superimposed, overlapping spots appeared *white* in the resulting image. (**b**) The images from lysate samples before (control, –) and after (+) treatment with Src kinase inhibitor I are shown in the *left-hand* and *center panels*, respectively. The two images are superimposed in the *right-hand panel*. The images obtained by detection using the lysates treated with or without the inhibitor are represented by *green* and *magenta colors*, respectively. *Green* and *magenta colors* indicate increases and decreases in kinome activities after treatment with the inhibitor, respectively. Reproduced from ref. 32 with permission from WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

4 Notes

1. All reagents and solvents used are purchased at the highest commercial quality available and used without further purification. All aqueous solutions are prepared by using deionized and distilled water.
2. Hydrochloric acid (HCl) is dangerously irritating to the skin, eyes, and mucous membranes. When handing this chemical, work in a chemical fume hood and wear gloves, eye protection, and a mask.
3. Sodium hydroxide (NaOH) is dangerously irritating to the skin and eyes. When handing this chemical, wear gloves and eye protection.
4. Each CelluSpots kinase substrate array [Tyr Kinase Substrate I (YKS-I) or Se/Thr Kinase Substrates II (STKS-II)] consists of 384 peptide–cellulose conjugate spots printed in duplicate on a glass slide (76 × 26 mm). Those include four control spots consisting of two types of phosphopeptide, one with phosphorylated Ser/Thr residues (spot #M22) and one with a phosphorylated Tyr residue (spot #O2). The three-dimensional

layer of the conjugates holds up to 1000 times more peptide in a given area compared with monolayer deposition, resulting in high sensitivity of detection.

5. The manufacturer's instructions state that the arrays are stable for at least 3 months under these conditions. The array surface is stable under incubation conditions, but is sensitive to mechanical stress and should not be touched or wiped.
6. The optimal concentration of cell lysates to achieve sufficient detection should be determined.
7. Zinc(II) nitrate solution [10 mM $Zn(NO_3)_2 \cdot 6H_2O$ in distilled water] is a suitable substitute. Because $ZnCl_2$ and $Zn(NO_3)_2$ are deliquescent salts, the solutions should be prepared by using fresh products from newly opened bottles. Aqueous solutions of $ZnCl_2$ or $Zn(NO_3)_2$ are stable for at least 6 months.
8. The commercially available Solution B and Solution D are used as received.
9. Phos-tag Biotin in Solution B (10 μL) is present in a large excess compared with HRP-conjugated streptavidin in Solution D (1 μL). We obtained the same result by using a smaller amount of Phos-tag Biotin (e.g., 1 μL of Solution B). The volume of Solution B (1–10 μL) can be adjusted appropriately to obtain the required sensitivity or to reduce expense. If the volume of Solution B is decreased from 10 to 1 μL, there is no need to adjust the volumes of the other solutions.
10. If the blocking treatment is omitted, many false-positive ECL signals can be detected (*see* Fig. 2a, left-hand panel). As mentioned above, the CelluSpots are microarrays of cellulose-conjugated peptides spotted on the planar surface of a glass slide to form a three-dimensional layer. Presumably, nonspecific interactions between the cellulose and HRP-conjugated streptavidin are responsible for the false-positive signals.
11. The peptide-spotted area should be totally covered by the Kinome Reaction Buffer. When the solution is dropped by a pipette, the tip of pipette should not contact the surface of the array.
12. Fluorescence-based macroarray methods using Phos-tag Biotin have also been developed by other groups [36–38].

Acknowledgments

This work was supported in part by KAKENHI Grants (24590050, 25293005, 25560417, 25117718, and 26460036) and by a research grant from the Takeda Science Foundation.

References

1. Hunter T (2000) Signaling: 2000 and beyond. Cell 100:113–127
2. Olsen JV, Blagoev B, Gnad F et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648
3. Ubersax JA, Ferrell JE Jr (2007) Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 8:530–541
4. Manning G, Whyte DB, Martinez R et al (2002) The protein kinase complement of the human genome. Science 298:1912–1934
5. Brognard J, Hunter T (2011) Protein kinase signaling networks in cancer. Curr Opin Genet Dev 21:4–11
6. Patterson H, Nibbs R, McInnes I et al (2014) Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. Clin Exp Immunol 176:1–10
7. Martin L, Latypova X, Terro F (2011) Post-translational modifications of tau protein: implications for Alzheimer's disease. Neurochem Int 58:458–471
8. Ishida A, Kameshita I, Sueyoshi N et al (2007) Recent advances in technologies for analyzing protein kinases. J Pharmacol Sci 103:5–11
9. Kosako H, Nagano K (2011) Quantitative phosphoproteomics strategies for understanding protein kinase-mediated signal transduction pathways. Expert Rev Proteomics 8:81–94
10. Schutkowski M, Reineke U, Reimer U (2005) Peptide arrays for kinase profiling. ChemBioChem 6:513–521
11. Arsenault R, Griebel P, Napper S (2011) Peptide arrays for kinome analysis: new opportunities and remaining challenges. Proteomics 11:4595–4609
12. Yamamoto T, Mori T, Katayama Y (2014) Microarray technologies for intracellular kinome analysis. Curr Med Chem 21:2542–2552 doi:10.2174/0929867321666131212154153
13. Kinoshita E, Takahashi M, Takeda H et al (2004) Recognition of phosphate monoester dianion by an alkoxide-bridged dinuclear zinc(II) complex. Dalton Trans 1189–1193
14. Kinoshita E, Yamada A, Takeda H et al (2005) Novel immobilized zinc(II) affinity chromatography for phosphopeptides and phosphorylated proteins. J Sep Sci 28:155–162
15. Kinoshita-Kikuta E, Kinoshita E, Yamada A et al (2006) Enrichment of phosphorylated proteins from cell lysate using a novel phosphate-affinity chromatography at physiological pH. Proteomics 6:5088–5095
16. Kinoshita-Kikuta E, Kinoshita E, Koike T (2009) Phos-tag beads as an immunoblotting enhancer for selective detection of phosphoproteins in cell lysates. Anal Biochem 389:83–85
17. Kinoshita-Kikuta E, Yamada A, Inoue C et al (2010) A novel phosphate-affinity bead with immobilized Phos-tag for separation and enrichment of phosphopeptides and phosphoproteins. J Integr OMICS 1:157–169
18. Tsunehiro M, Meki Y, Matsuoka K et al (2013) A Phos-tag-based magnetic-bead method for rapid and selective separation of phosphorylated biomolecules. J Chromatogr B Anal Technol Biomed Life Sci 925:86–94
19. Kinoshita-Kikuta E, Aoki Y, Kinoshita E et al (2007) Label-free kinase profiling using phosphate affinity polyacrylamide gel electrophoresis. Mol Cell Proteomics 6:356–366
20. Kinoshita E, Kinoshita-Kikuta E, Matsubara M et al (2008) Separation of phosphoprotein isotypes having the same number of phosphate groups using phosphate-affinity SDS-PAGE. Proteomics 8:2994–3003
21. Kinoshita E, Kinoshita-Kikuta E, Matsubara M et al (2009) Two-dimensional phosphate-affinity gel electrophoresis for the analysis of phosphoprotein isotypes. Electrophoresis 30:550–559
22. Kinoshita E, Kinoshita-Kikuta E, Ujihara H et al (2009) Mobility shift detection of phosphorylation on large proteins using a Phos-tag SDS-PAGE gel strengthened with agarose. Proteomics 9:4098–4101
23. Kinoshita E, Kinoshita-Kikuta E, Koike T (2009) Separation and detection of large phosphoproteins using Phos-tag SDS-PAGE. Nat Protoc 4:1513–1521
24. Kinoshita E, Kinoshita-Kikuta E (2011) Improved Phos-tag SDS-PAGE under neutral pH conditions for advanced protein phosphorylation profiling. Proteomics 11:319–323
25. Kinoshita E, Kinoshita-Kikuta E, Koike T (2012) Phos-tag SDS-PAGE systems for phosphorylation profiling of proteins with a wide range of molecular masses under neutral pH conditions. Proteomics 12:192–202
26. Kinoshita-Kikuta E, Kinoshita E, Koike T (2012) Separation and identification of four distinct serine-phosphorylation states of ovalbumin by Phos-tag affinity electrophoresis. Electrophoresis 33:849–855
27. Kinoshita-Kikuta E, Kinoshita E, Koike T (2012) A laborsaving, timesaving, and more reliable strategy for separation of low-molecular-mass phosphoproteins in Phos-tag affinity electrophoresis. Int J Chem (Mumbai, India) 4(5):1–8

28. Kinoshita E, Kinoshita-Kikuta E, Shiba A et al (2014) Profiling of protein thiophosphorylation by Phos-tag affinity electrophoresis: evaluation of adenosine 5′-*O*-(3-thiotriphosphate) as a phosphoryl donor in protein kinase reactions. Proteomics 14:668–679
29. Kinoshita E, Kinoshita-Kikuta E, Takiyama K et al (2006) Phosphate-binding tag, a new tool to visualize phosphorylated proteins. Mol Cell Proteomics 5:749–757
30. Inamori K, Kyo M, Nishiya Y et al (2005) Detection and quantification of on-chip phosphorylated peptides by surface plasmon resonance imaging techniques using a phosphate capture molecule. Anal Chem 77:3979–3985
31. Nakanishi T, Ando E, Furuta M et al (2007) Identification on membrane and characterization of phosphoproteins using an alkoxide-bridged dinuclear metal complex as a phosphate binding tag molecule. J Biomol Tech 18:278–286
32. Kinoshita E, Kinoshita-Kikuta E, Sugiyama Y et al (2012) Highly sensitive detection of protein phosphorylation by using improved Phos-tag Biotin. Proteomics 12:932–937
33. Kinoshita E, Kinoshita-Kikuta E, Koike T (2013) Highly sensitive detection method for protein phosphorylation and its development as a new technique toward phosphoproteomics. Seikagaku 85:447–455 (in Japanese)
34. Kinoshita E, Kinoshita-Kikuta E, Koike T (2013) Phos-tag-based microarray techniques advance phosphoproteomics. J Proteomics Bioinf S6:008
35. Kinoshita E, Kinoshita-Kikuta E, Koike T (2013) Sandwich assay for phosphorylation of protein multiplexes by using antibodies and Phos-tag. Anal Biochem 438:104–106
36. Han X, Sonoda T, Mori T et al (2010) Protein kinase substrate profiling with a high-density peptide microarray. Comb Chem High Throughput Screen 13:777–789
37. Han X, Katayama Y (2010) A peptide microarray for detecting protein kinase activity in cell lysates. Methods Mol Biol 669:183–194
38. Ikeda H, Kamimoto J, Yamamoto T et al (2013) A peptide microarray fabricated on a non-fouling phosphatidylcholine-polymer-coated surface for a high-fidelity analysis of a cellular kinome. Curr Med Chem 20:4419–4425

Chapter 3

Phosphopeptide Enrichment by Covalent Chromatography After Solid Phase Derivatization of Protein Digests on Reversed Phase Supports

Heinz Nika, Ruth Hogue Angeletti, and David H. Hawke

Abstract

The isolation of the phosphopeptide constituents from phosphoprotein digests is prerequisite to facilitate the mass spectrometric characterization of phosphorylation events. Here, we describe a chemical proteomics approach which combines solid phase derivatization of phosphoprotein digests with phosphopeptide enrichment by covalent chromatography. The use of the solid phase support for derivatization ensures for speed and completeness of reactions. The isolates proved highly suitable for mapping of the sites of phosphorylation by collisionally induced dissociation (CID). The method combines robustness with simplicity of operation using equipment available in biological laboratories, and may be readily extended to map the sites of O-glycosylation.

Key words Mass spectrometry, β-elimination/Michael addition, Reversed phase support, Multi-step solid-phase derivatization, Phosphopeptide enrichment, Covalent chromatography, Phosphorylation site determination

1 Introduction

Protein phosphorylation is recognized as a critical event in modulation of cellular processes including cellular signaling, cell cycle progression and differentiation [1].

Therefore, the determination of phosphorylation sites by mass spectrometry (MS) has become increasingly important. However, these efforts are challenged by the ionization inefficiency of the phosphopeptides, their low stoichiometry and the limited information content of the tandem MS (MS/MS) spectra due to neutral loss of the phosphate group upon collision induced dissociation (CID). Immobilized metal ion affinity chromatography (IMAC) and titanium dioxide (TiO_2) chromatography are frequently used techniques to address these problems by phosphopeptide enrichment [2, 3]. In many instances these methods had been inefficient due to co-adsorption of non-phosphorylated (acidic) peptides.

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_3,

TiO_2 chromatography has been shown to poorly enrich phosphopeptides derived from basophilic kinase substrates. Although the above strategies afford improved phosphopeptide detection, issues still persist with regard to intrinsic instability of phosphate groups often rendering site mapping subject to ambiguity under the conditions of CID.

β-elimination of phosphate from serine and threonine residues coupled with Michael addition (BEMAD) provides for a chemical strategy to enrich phosphopeptides from unfractionated protein digests. This approach has been successfully adopted by several laboratories to map phosphorylation events in isolated proteins and on the proteome-wide scale [4–9] and has also been exploited to enrich for proteolytic fragments bearing O-linked β-N-acetylglucosamine (O-GlcNAc) [10]. In this strategy, dithiothreitol (DTT) was used as nucleophile in the Michael addition reaction. The thiol adducts were then captured by thiol-disulfide interchange on activated Thiol Sepharose, and reductively released from the affinity support; a procedure based on the concept of covalent chromatography [5]. Differentiation between phosphorylation and O-glycosylation was afforded by enzymatic dephosphorylation which renders the phosphopeptide component in the mixture insensitive to enrichment [10]. The nucleophilic substitution of the phosphorylated (glycosylated) residues precluded neutral loss of phosphate (glycan); the preferential pathway of fragmentation by CID. In consequence, the isolates sequenced noticeably more informatively than their native counterparts facilitating phosphorylation site determination. However, the application of the method to the characterization of the phosphoproteome of human whole saliva revealed that the enriched fractions contained mostly singly phosphorylated peptides [9]. We encountered the same complication in attempts to isolate the DTT adduct of the tetraphosphorylated peptide of β-casein [11]. We found that cross-link formation between the adjacent phosphosites through intramolecular Michael addition rendered the peptide inert to enrichment by covalent chromatography. This observation strongly suggests that this class of hyper-phosphorylated peptides, accounting for ~37 % of phosphopeptides in eukaryotic proteins, is in general excluded from isolation by the dithiol-based affinity enrichment approach [12]. Notably, the phosphoramidate chemistry (PAC) used for large-scale phosphopeptide isolation from digests of *Drosophila melanogaster* Kc167 cells also exhibited a strong bias against the selection of multiply phosphorylated peptides [13]. In addition, this method regenerates the original phosphopeptides rendering site mapping subject to ambiguity under the conditions of CID. These shortcomings prompted us as to devise an alternative strategy and the method that evolved from these efforts is the subject of the protocol [11].

The method's work-flow and a schematic depiction of the structure of the analyte and its derivatives are illustrated in Fig. 1a, b, respectively. In this protocol the phosphoprotein digest is extracted on a ZipTip$_{C18}$ pipette tip (SPE) and submitted in situ to sequential performic acid oxidation, acetylation and BEMAD using 2-aminoethanethiol as nucleophile. The amine-protection step renders the amino groups of the Michael addition products (i.e. the *N*-acetyl *S*-2-aminoethyl-/β-methyl-*S*-2-aminoethylcysteine derivatives) as the sole targets for subsequent acylation with sulfosuccinimidyl-2-(biotinamido)-ethyl-1, 3-dithiopropionate (Sulfo-NHS-SS-Biotin). In this sequence of solid phase reactions, denoted as N-thiolation, the biotinylated digest is then exposed to hydroxylamine for reversal of hydroxyl group acylation followed by reductive release of the disulfide-linked biotinamido moiety from the conjugates. The Michael addition derivatives, selectively thiolated

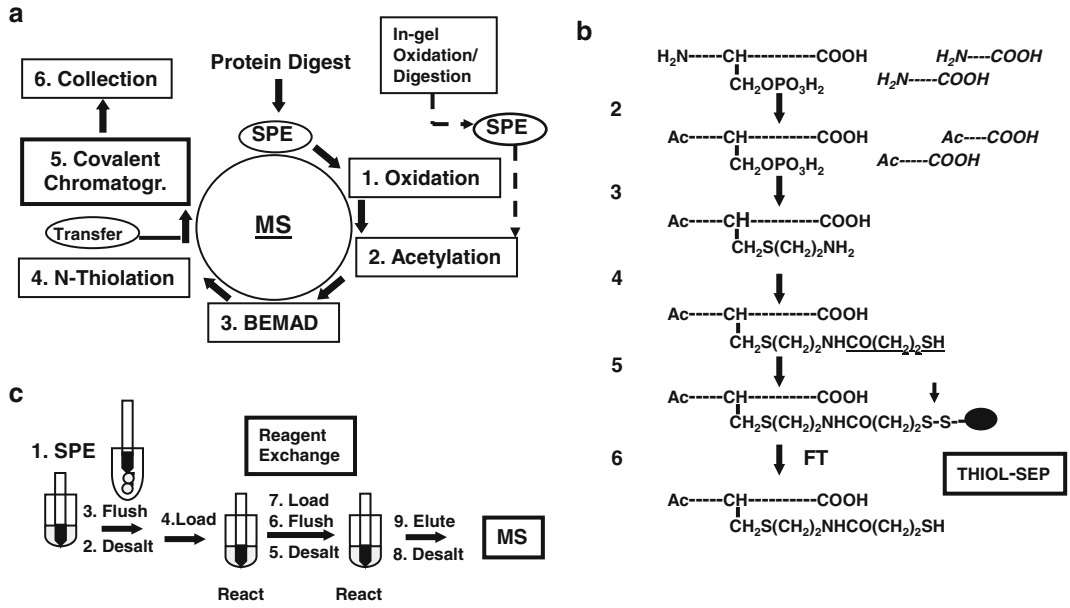


Fig. 1 (**a**) Work-flow for Phosphopeptide Enrichment. The reaction steps are numbered and highlighted with boxes. The digest is extracted on a ZipTip$_{C18}$ pipette tip (SPE), carried through the sample preparation scheme and submitted to covalent chromatography. Non-bound material is removed by a solvent wash. The disulfide-bonded phosphopeptide analogs are reductively released from the affinity support and collected in the flow-through fraction. BEMAD designates β-elimination with Michael addition. Flow path modification accommodating digests recovered from in-gel oxidized protein is indicated by *stippled arrows*. (**b**) Schematic representation of the analyte and its derivatives. The affinity tag is *underlined*. Ac and THIOL-SEP denote acetyl group and activate-Thiol Sepharose, respectively. Non-phosphorylated peptides are highlighted in *italics*. (**c**) Schematic representation of solid phase sample handling steps. After solid phase extraction (SPE) the analyte is desalted. Then the ZipTip is briefly flushed with reagent, loaded with reagent and incubated while immersed in reagent. In this manner reagents are exchanged between the reaction steps in situ eliminating intermittent sample transfer. The reaction cycle is concluded by a solvent wash prior to product elution

in this manner, are then enriched by covalent chromatography, reductively released from the affinity support and collected in the flow-through fraction. Co-isolation of cysteinyl peptides is prevented by the initial reaction step which converts cysteine to its sulfonic acid analog, thereby making cysteinyl peptides insensitive toward BEMAD. Alternatively, gel-separated protein can be oxidized and digested in the gel matrix [11]. In this way the reactivity of cysteine toward BEMAD is arrested prior to adsorption of the digest on the reversed phase support.

Figure 1c depicts the handling steps employed to carry the digest through the sample preparation method. After SPE the immobilized digest is desalted, briefly flushed with reagent which is then loaded onto the reversed phase support. In the process of adsorption the analyte is concentrated on the solid phase typically by a factor of 100 or more relative to solutions of the same original concentration [14]. As a result, chemistries proceed on the solid phase at higher efficiency and faster kinetics than in solution providing a significant advantage when dilute samples known to react poorly in solution are processed by this technique. In the process of serial derivatization reagents are exchanged in situ thereby eliminating intermittent sample transfer. In consequence, the derivatives are carried through the multi-step sample preparation scheme with minimal loss; a task that in general has met with moderate success when sequential derivatization is carried out in the solution phase; the predominant reaction format in proteomics studies [15]. It is noteworthy that modern analytical sample preparation techniques for the reasons outlined above are almost exclusively based on solid phase derivatization and have found widespread use for automated, high-throughput trace analysis of bioorganic compounds in toxicology, environmental, and pharmaceutical studies [14].

2 Materials

Prepare all solutions using ultrapure water with 18 MΩ cm resistivity at 25 °C. Unless stated otherwise prepare reagent solutions fresh for daily use. Reagents should be of analytical grade.

2.1 Gel-Electrophoresis and In-Gel Protein Digestion Reagent/Solvent Setup

1. 10 % gels of 1 mm thickness (Criterion Tris–HCl Precast gel, 133×87 mm, 18 wells, Bio-Rad, Hercules, CA) or alternatively Criterion Tris–HCl Precast 10–20 % gradient gels (same dimensions as the continuous gels).
2. GelCode Blue Stain Reagent (Pierce Corp., Rockford, IL), a Coomassie blue-based reagent.
3. Gel Destaining Solution: 25 mM ammonium bicarbonate, 50 % aqueous acetonitrile. Prepare a 0.25 M ammonium bicarbonate stock solution. Dissolve 179.6 mg of ammonium bicar-

bonate in 10 mL water. Mix 1 mL 0.25 M stock solution with 5 mL acetonitrile and 4 mL of water. Store the Destaining Solution at 4 °C.

4. Disulfide Reductant: 2 mM tris [2-carboxyethyl] phosphine hydrochloride (TCEP), 25 mM ammonium bicarbonate. Prepare a 25 mM ammonium bicarbonate solution by diluting the 0.25 M stock solution with water in a 1:10 ratio. Add 8 μL of the 0.5 M TCEP stock solution (Bond Breaker TCEP solution, Pierce Corp, Rockford, IL) to 1992 μL 25 mM ammonium bicarbonate. Store reagent at 4 °C.
5. Digestion Buffer: 25 mM ammonium bicarbonate, 0.01 % of *N*-octyl glucoside (OGS) (Roche Diagnostics, Indianapolis, IN). Dissolve 20 mg OGS in 1 mL 25 mM ammonium bicarbonate. Add 10 μL of the 2 % OGS stock solution to 1990 μL 25 mM ammonium bicarbonate. Store buffer at 4 °C.
6. Enzyme Solution: Sequencing grade modified trypsin (Promega, Madison, WI), 25 mM ammonium carbonate, 0.01 % OGS. Dissolve 25 μg of enzyme in 2 mL digestion buffer (12.5 ng/μL). Store enzyme solution in 100 μL aliquots at −20 °C. The enzyme remains active under this condition for several months.
7. Peptide Extraction Solvent: 0.1 % TFA. Prepare a 1 % aqueous TFA stock solution by mixing 10 μL of neat TFA with 990 μL water. Prepare 0.1 % TFA by diluting the stock solution with water in 1:10 ratio. Store solvent at 4 °C.

2.2 Solid Phase Derivatization Reagent/Solvent Setup

1. Performic Acid Oxidation Reagent: 0.3 % performic acid. Prepare reagent by mixing 30 % aqueous hydrogen peroxide with 96 % formic acid (5:95, v/v). Leave the reagent at room temperature for at least 30 min (*see* **Note 1**). Dilute with water at a 1:5 ratio to a final concentration of 0.3 % (pH 3.1). Store reagent at 4 °C for up to 1 h.
2. Acetylation Reagent: 20 mM of *N*-hydroxy sulfosuccinimide ester of acetic acid, 20 mM sodium phosphate (pH 8.0). Prepare a 0.25 M sodium phosphate buffer stock solution. Dissolve 3.45 g of $NaH_2PO_4 \cdot H_2O$ in 50 ml water (0.5 M Solution A). Dissolve 9.95 g of $Na_2HPO_4 \cdot 12H_2O$ in 50 ml water (0.5 M Solution B). Complete dissolution may require heating and vigorous stirring. Mix 23.7 ml Solution B with 1.3 ml Solution A and add 25 ml of water. Store the 0.25 M phosphate buffer stock solution at 4 °C. Mix 0.8 ml of the stock solution with 9.2 ml of water to give a final concentration of 20 mM sodium phosphate. Dissolve 5.2 mg Sulfo-NHS acetate (Pierce Corp., Rockford, IL) in 1 mL 20 mM sodium phosphate (*see* **Note 2**).

3. Concurrent BEMAD Reagent: 66 mM barium hydroxide, 33 mM 2-aminoethanethiol hydrochloride (pH 12.3). Prepare a 100 mM barium hydroxide stock solution by dissolving 31.54 mg of barium hydroxide octahydrate (Sigma-Aldrich, Milwaukee, WI) in 1 mL water (*see* **Note 3**). Prepare a 100 mM 2-aminoethanethiol hydrochloride stock solution by dissolving 11.36 mg of 2-aminoethanethiol hydrochloride (Sigma-Aldrich, Milwaukee, WI) in 1 mL water. Mix the barium hydroxide solution with the 2-aminoethanethiol hydrochloride solution in a ratio of 2:1. Store reagent at room temperature for daily use.
4. Consecutive BEMAD Reagents:
   (a) β-elimination Base: 50 mM barium hydroxide. Dilute 100 mM barium hydroxide with water in a 1:1 ratio. Store reagent at room temperature for daily use.
   (b) Michael Addition Reagent: 100 mM 2-aminoethanethiol hydrochloride, 75 mM barium hydroxide (pH 10.6). Mix 100 mM barium hydroxide with 400 mM 2-aminoethanethiol hydrochloride (45.44 mg/mL) in a 3:1 ratio. Store reagent at room temperature for daily use.
5. Biotinylation Reagent: 20 mM Sulfo-NHS-SS-Biotin (Pierce Corp., Rockford, IL), 20 mM sodium phosphate (pH 8.0). Dissolve 6 mg Sulfo-NHS-SS-Biotin in 1 mL 20 mM sodium phosphate (*see* **Note 4**).
6. O-deacylation Reagent: 2 % hydroxylamine hydrochloride (Pierce Corp., Rockford, IL), 1 M sodium carbonate (pH 9.4). Prepare a 1 M sodium carbonate solution by dissolving 105.9 mg sodium carbonate in 1 mL water. Dissolve 20 mg hydroxylamine hydrochloride in 1 mL 1 M sodium carbonate. Store reagent at room temperature for daily use.
7. Disulfide Reductant: 5 mM TCEP, 20 mM sodium phosphate (pH 8.0). Add 20 µL of the 0.5 M TCEP stock solution to 980 µL of 20 mM sodium phosphate (pH 8.0). Store reagent at room temperature for daily use.
8. Sample Cleanup Solution: 2 mM ethylenediaminetetraacetic acid (EDTA). Dissolve 7.3 mg EDTA in 10 mL water. Store the solution at 4 °C.
9. Sample Cleanup Solvent: 0.1 % TFA. Prepare a 1 % TFA stock solution by mixing 10 µL of neat TFA with 990 µL water. Prepare 0.1 % TFA by diluting the stock solution with water in 1:10 ratio. Store solvent at 4 °C.
10. Sample Eluate: 50 % aqueous acetonitrile, 0.1 % TFA, 0.01 % OGS. Mix 390 µL water with 500 µL acetonitrile and 100 µL 1 % TFA. Add 10 µL 2 % OGS. Store eluate at 4 °C.

2.3 Affinity Purification Reagent/ Solvent Setup

1. Activated-Thiol Sepharose 4B Gel: Weigh 1 g of activated-Thiol Sepharose 4B (GE Healthcare, Piscataway, NJ) and place the medium into a scintillation glass vial or equivalent. Add 10 ml water. Mix gently. The resin swells within 10 min and should give a ~4 ml settled medium. Wash medium in an appropriate glass filter funnel by vacuum filtration with a total of 150 ml of water added in 15 ml aliquots. Suspend medium in 10 ml of 10 % aqueous ethanol with gentle agitation, transfer medium to a 10 ml scintillation glass vial or equivalent. Store medium at 4 °C for at least 2 weeks (*see* **Note 5**).
2. Affinity Coupling Buffer: 50 mM sodium phosphate, 2 mM EDTA (pH 8.0). Add to a 15 mL Falcon tube 2 mL 0.25 M sodium phosphate stock solution and 8 mL water. Add 7.3 mg EDTA. Vortex vigorously to facilitate dissolution. Store the Coupling Buffer at 4 °C.
3. Disulfide Reductant: 50 mM sodium phosphate, 5 mM TCEP (pH 8.0). Add to a 15 mL Falcon tube 0.5 mL of the 0.25 M sodium phosphate stock solution, 1.975 mL water and 25 μL 0.5 M TCEP. Store the reductant at 4 °C.
4. Resin Wash Solvent A: 60 % acetonitrile, 0.1 %. TFA. Prepare a 10 % TFA stock solution by adding 0.1 mL neat TFA to 0.9 mL water. Mix 6 ml acetonitrile with 3.9 mL water and 0.1 ml 10 % TFA. Store solvent at 4 °C.
5. Resin Wash Solvent B: 80 % acetonitrile, 0.1 % TFA. Mix 8 mL of acetonitrile with 1.9 mL water and 0.1 mL 10 % TFA. Store solvent at 4 °C.
6. α-cyano-4-hyroxycinnamic acid (Agilent Technologies, Palo Alto, CA), 0.1 % TFA. Add 10 μL 1 % aqueous TFA to 90 μL matrix.

2.4 Test Analysis and Diagnostic Test

1. Bovine α-S1 and β-casein.
2. Test peptides: DAM1 phosphopeptide SFVLNPTNIGMp SKSSQGHVTK (AnaSpec, San Jose, CA); angiotensin I peptide DRVYIHPFHL (Sigma-Aldrich, St. Louis, MO); cysteinyl peptide Somatostatin CKNFFWKT, *m/z* 1073.2 (Sigma-Aldrich, St. Louis, MO).

2.5 Equipment

1. Gilson pipettor P20, model Pipetman classic, 2–20 μL displacement or equivalent.
2. Bench top centrifuge Eppendorf 5415 D or equivalent.
3. Rotary mixer.
4. Modular block heater.
5. Thermomixer.
6. Savant SpeedVac concentrator.

7. MALDI-TOF mass spectrometer (Voyager DE STR, Applied Biosystems, CA) or equivalent.
8. MALDI-TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, CA) or equivalent.
9. Spectra were analyzed using Data Explorer (Applied Biosystems, version 4.8).

2.6 Other Materials

1. ZipTip$_{C18}$ pipette tips (0.6 µL bed volume, Millipore Corp., Billerica, MA) or with ZipTip$_{µ\text{-}C18}$ pipette tips (0.2 µL bed volume) for solid phase reactions.
2. Eppendorf LoBind microcentrifuge tubes (0.5, 1.5, and 2 mL).
3. Eppendorf pipette tips.
4. Spin Columns-Screw Caps, 0.8 mL internal volume (Pierce Corp, Rockford, IL).

3 Methods

3.1 In-Gel Digestion

After sodium dodecylsulfate-polyacrylamide electrophoresis (SDS-PAGE) and gel staining (*see* **Note 6**), carry out all manipulations in a laminar flow-vented hood or equivalent. Wear gloves at all times to prevent keratin contamination.

1. Run your protein sample on a gel.
2. Stain using Colloidal Coomassie using instructions supplied by the manufacturer (most dye-stains will work).
3. Excise protein band from the stained gel using a clean scalpel.
4. Cut band into ~1 × 1 mm wide pieces and transfer to a 0.5 mL centrifuge tube.
5. Set thermomixer to an agitation speed of 500 rpm.
6. Incubate gel pieces twice with 200 µl of the Destaining Solution for 30 min at 37 °C; discard supernatant.
7. Dry gel band briefly in a SpeedVac.
8. Add 100 µL of the reductant and incubate for 15 min at 37 °C (*see* **Note 7**).
9. Remove supernatant.
10. Dehydrate the gel pieces for 10 min in 100 µL of acetonitrile. Discard supernatant and briefly dry the gel in a SpeedVac.
11. Rehydrate gel pieces at room temperature in 20 µL Promega-modified trypsin solution (12.5 ng/µL). Discard the remainder of the enzyme solution.

12. After 30 min add 40 µL 25 mM ammonium bicarbonate and incubate for 18 h at 37 °C.
13. After incubation, add 50 µl 0.1 % TFA and collect supernatant in a 1.5 mL centrifuge tube (Extract 1). Prior to collection of the extract place 35 µL of 0.1 % TFA into the centrifuge tube. Mark the liquid level. Empty the tube.
14. Add 50 µl 0.1 % TFA to the gel pieces and incubate for 30 min at 37 °C (Extract 2).
15. Transfer Extract 2 to the collection tube.
16. Reduce the extracts in volume to 35 µL by SpeedVac evaporation (*see* **Note 8**).
17. Add 5 µL of 10 % TFA. Proceed with sample binding.

3.2 Sample Binding to the Reversed Phase Support

1. Wet the ZipTip$_{C18}$ pipette tip six times with 10 µL of methanol followed by six 10 µL washes with 0.1 % TFA according to the manufacturer's instructions. Critical! Do not allow the ZipTip to dry before sample loading. Should this inadvertently occur, condition a new ZipTip and proceed immediately with the next step. Aspirate the acidified sample in 10 µL aliquots onto the ZipTip and dispense into a 0.5 ml microfuge tube. Transfer the partially stripped peptide solution back in this manner into the original collection tube. Repeat this alternating enrichment cycle five times to maximize peptide recovery. Then wash the ZipTip with 50 µL 0.1 % TFA by passing the solvent in ten aliquots over the resin. Critical! After the desalting step do not allow the ZipTip to de-wet. Proceed immediately to the performic acid oxidation step or aspirate 10 µL 0.1 % TFA onto the ZipTip for the purpose of temporary storage (*see* **Note 9**). Prepare model peptide solutions at a concentration of 2 pmol/µL in 1 % aqueous TFA, 0.01 % OGS. Transfer the solutions in five10 µL aliquots to 0.5 mL centrifuge tubes. Bind peptides to the support using up to ten sample aspiration/dispense cycles. Wash ZipTips with 50 µL 0.1 % TFA. Proceed with the experiments described under Subheading 3.6 or store the ZipTips temporarily in 0.1 % TFA until use.

3.3 Solid Phase Derivatization

With the exceptions noted below the ZipTip$_{C18}$ pipette tip is intermittently desalted. Then the ZipTip is flushed twice with reagents. After this resin conditioning step, the reagents are loaded onto the support from the 60 µL that had been placed into 0.5 mL microcentrifuge tubes. Critical! Leave the ZipTip immersed in reagents during the incubations. Do not allow the ZipTip to dry during intermittent sample handling. If the ZipTip is inadvertently allowed to de-wet prior to reagent loading repeat the resin conditioning step. Agitation during incubations is not required. Use the modular block heater or equivalent device.

1. Flush the ZipTip to waste with 10 µL of the oxidant. Repeat this step. Load 10 µL of the oxidant onto the support. Allow reaction to proceed for 1 h at 4 °C. Halt the reaction by passing 100 µL of 0.1 % TFA in ten aliquots over the resin.
2. Condition the desalted ZipTip twice with 10 µL of the sulfo-NHS acetate reagent. Load 10 µL of reagent onto the support. Allow reaction to proceed for 20 min at 55 °C.
3. Flush the ZipTip twice with 10 µL of the BEMAD reagent mixture. Load 10 µL of the reagent mixture onto the support and allow the concurrent reaction to proceed for 1 h at 37 °C. Wash the ZipTip ten times with 10 µL 0.1 % TFA. Implement the consecutive BEMAD reaction mode as follows: (*see* **Note 10**). Condition the ZipTip twice with 10 µL β-elimination Base. Aspirate 10 µL reagent onto the ZipTip. Incubate for 30 min at 55 °C. Flush the ZipTip twice with 10 µL Michael Addition Reagent. Aspirate 10 µL of reagent onto the ZipTip and allow the addition reaction to proceed for 2 h at 55 °C. Desalt the ZipTip by passing 100 µL of 0.1 % TFA in 10 µL aliquots over the resin. *Pause point*. The ZipTip may be stored overnight in 0.1 % TFA at −21 °C.
4. Condition the desalted ZipTip twice with 10 µL of the sulfo-NHS-SS-Biotin Reagent. Load 10 µL of reagent onto the support. Incubate for 30 min at room temperature.
5. Flush the ZipTip twice with 10 µL of the hydroxylamine solution. Load 10 µL of reagent onto the ZipTip and allow the reaction to proceed for 15 min at 37 °C. Wash the resin ten times with 10 µL of 0.1 % TFA and then five times with 10 µL of water. *Pause point*. The ZipTip may be stored overnight in 0.1 % TFA at −21 °C.
6. Condition the desalted ZipTip twice with 10 µL of the reductant. Aspirate 10 µL of the reductant onto the support. Incubate for 15 min at 37 °C. Wash the ZipTip ten times with 10 µL 2 mM aqueous EDTA and then ten times with 10 µL of 0.1 % TFA. Critical! This stringent purification step is required for thorough removal the reductant. Residual TCEP would diminish the binding capacity of the affinity support. Store the ZipTip at 4 °C in 0.1 % TFA while proceeding to **step 9**, Subheading 3.4.

3.4 Covalent Chromatography

1. Place the spin column in a 2 ml centrifuge tube. Add 200 µL of acetonitrile and centrifuge at 209 × *g* for 1 min.
2. Let the activated-Thiol Sepharose gel, stored at 4 °C in 10 mL 10 % aqueous ethanol, assume room temperature.
3. Adjust the liquid level in the storage container (i.e., the scintillation tube) to one third of the height of the settled gel.

4. Swirl the container slightly to suspend the medium. Pipette 100 µl of the slurry into the spin column. To facilitate pipetting, shorten a 200 µL pipette tip by ~2 mm using a clean razor blade.
5. Centrifuge at 209 × *g* for 1 min.
6. Suspend the medium pellet in 200 µl of Coupling Buffer and centrifuge at 209 × *g* for 1 min.
7. Seal spin column with the plastic plug supplied by the manufacturer. Ensure that the plug is properly seated to avoid leaks.
8. Insert spin column into a 1.5 mL centrifuge tube.
9. Elute the digest from the ZipTip (set aside in **step 6**, Subheading 3.3) with 10 µL 50 % acetonitrile, 0.1 % TFA, 0.01 % OGS into a 0.5 mL centrifuge tube. Place 1 µL of the eluent onto the MALDI target. Immediately add 40 µl Coupling Buffer to the remainder. Mix. Checkpoint! Analyze the aliquot along with the material recovered in **steps 25** and **26**, Subheading 3.4 (*see* **Note 11**).
10. Transfer the mixture to the spin column. Swirl the spin column gently to suspend the medium pellet.
11. Attach screw cap and secure the spin column to the centrifuge tube with a piece of Parafilm.
12. End over end incubate the sample for 1 h at room temperature in the rotary mixer.
13. Remove plug from the spin column and centrifuge at 209 × *g* for 1 min. Mark the liquid level (~50 µL) on the collection tube.
14. Place 50 µl of Coupling Buffer into the spin column, suspend the pellet, and centrifuge at 209 × *g* for 1 min.
15. Place 50 µl 60 % aqueous acetonitrile, 0.1 % TFA into the spin column, suspend the pellet with agitation, and centrifuge at 209 × *g* for 1 min.
16. Place 50 µl 80 % aqueous acetonitrile, 0.1 % TFA into the spin column, suspend the pellet with agitation and centrifuge at 209 × *g* for 1 min (*see* **Note 12**).
17. Reduce the volume of the collected fractions by SpeedVac evaporation to the level indicated on the collection tube. Add 5 µL of 15 % TFA. Store sample in 0.1 % TFA at 4 °C.
18. Insert the spin column into a 1.5 mL centrifuge tube.
19. Place 50 µl Coupling Buffer into to the spin columns to neutralize residual acid. Centrifuge at 209 × *g* for 1 min; discard the flow-through.
20. Seal the spin column and suspend the medium pellet with 50 µl of the reductant. Attach screw cap. Place the spin column

into a 1.5 ml centrifuge tube and secure the spin column to the centrifuge tube with a piece of Parafilm.

21. With end-over-end mixing, incubate for 30 min at room temperature.
22. Unplug the spin column and centrifuge at 209 × *g* for 1 min. Mark the liquid level on the collection tube (~50 µL).
23. Wash the affinity resin consecutively with the organic solvents as described above.
24. Reduce volume of the collected fractions by SpeedVac evaporation to the level indicated on the collection tube (*see* **Note 13**). Add 5 µL of 15 % TFA.
25. Bind the material to a ZipTip$_{C18}$ pipette tip, pass 100 µL 0.1 % TFA in 10 µL portions over the resin and elute with 5–10 µL 50 % aqueous acetonitrile, 0.1 % TFA, 0.01 % OGS. Subject 1 µL of the eluate to MALDI MS analysis (*see* **Note 14**).
26. Concentrate the material (i.e. the non-phosphorylated peptides) kept in temporary storage (**step 17**, Subheading 3.4) on a ZipTip$_{C18}$ pipette tip. Process the ZipTip for MALDI MS analysis as described above.

3.5 Anticipated Results and Diagnostic tests

3.5.1 Application of Method to Model Protein

To exemplify the performance of the sample preparation/covalent chromatographic system an in-gel digest prepared from an equimolar mixture of bovine α-S1 and β-casein (25 pmol) was examined. The results from this experiment that demonstrate the successful implementation of the method are shown in Fig. 2a–c [11]. On the basis of the known protein sequences the ions at *m/z* 1770.3, *m/z* 2103.2, and *m/z* 3432.6 are recognized as the *N*-acetyl thiol derivatives of the α-S1 monophosphorylated fragment -^{121}VPQLEIVPNpSAEER^{134}, of the miscleavage product ^{119}YKVPQLEIVPNpSAEER^{134} and of the tetraphosphorylated fragment of β-casein-^{16}RELEELNVPGEIVEpSLpSpSpSEESITR40, respectively which proved impervious to enrichment by the dithiol-based affinity approach (*see* Fig. 2c). Ions which would indicate incomplete acetylation were not observed in the spectrum of the starting material (*see* Fig. 2a, arrow heads). The ions at *m/z* 2060.3 and at *m/z* 2390.4 matched in mass to the oxidized counterparts of fragments harboring methionine. Non-phosphorylated peptides were nearly exclusively found in the solvent wash providing evidence that these peptides were effectively depleted from the affinity support (*see* Fig. 2b). The data provide a general means for method trouble shooting. Tests proposed to survey the individual reaction steps in detail for potential complications are described below.

The results of the application of the method to a digest prepared by in solution digestion of 2 pmol α-S1 casein are shown in Fig. 2c, inset (*see* **Note 15**). The high quality of the spectrum produced from the isolate suggests that sub-picomole quantities of

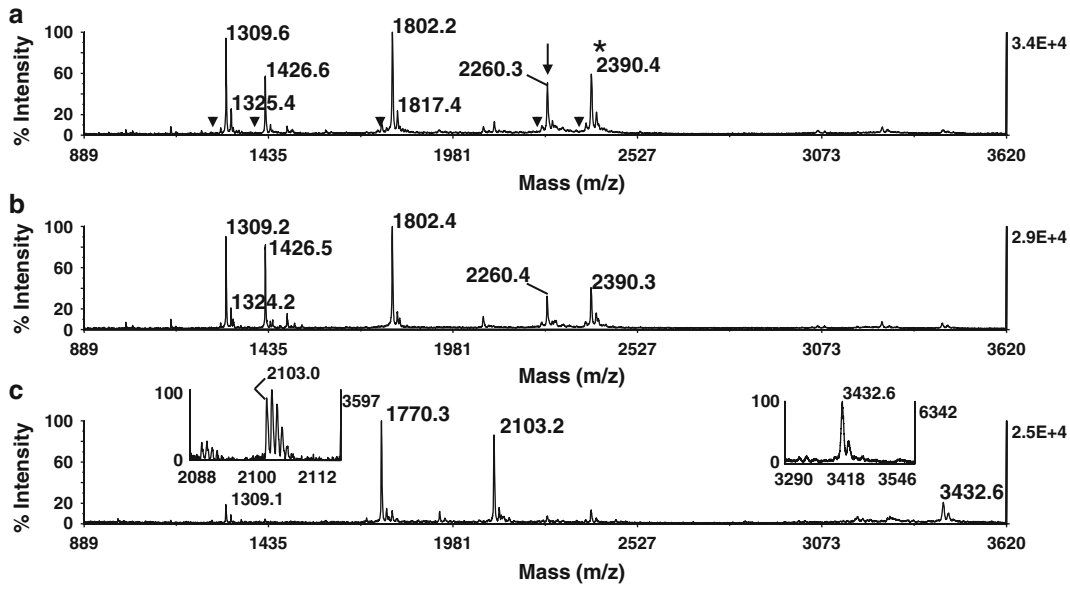


Fig. 2 Application of method to a tryptic in-gel digest prepared from equimolar mixture of 25 pmol of bovine α-S1 and β-casein. MALDI MS spectra of (**a**) digest after preconditioning reactions; (**b**) solvent wash; (**c**) reductively released fraction. The phosphopeptides are recognized at *m/z* 1770.3, *m/z* 2103.2, and *m/z* 3432.6 as the α-S1 casein fragments ^{121}VPQLEIVPpSAEER^{134} and ^{119}YKVPQLEIVPNpSAEER^{134} and the β-casein fragment ^{16}RELEELNVPGEIVEpSLpSpSpSEESITR^{40}, respectively. The fragment of β-casein and of α-S1casein targeted by oxidation is denoted by *arrow* and *asterisk*, respectively. *Inset* in (**c**), *right margin* shows expanded section of spectrum containing the tetraphosphorylated peptide. ~2 pmol of digest were applied to the target. *Inset* in (**c**), *left margin* shows expanded section of spectrum containing the α-S1 miscleavage product selected from 2 pmol of α-S1 digest. ~200 fmol of isolate were applied to the target

digest should be amenable to the chemical approach. This precept is supported by the finding that femtomole amounts of model peptides can be successfully carried through the reaction schemes [11]. See below.

3.5.2 Phosphorylation Site Determination by MALDI-TOF/ TOF MS

Phosphorylation site determination by CID in native phosphopeptides often fails due to lack of informative sequence information because gas-phase cleavage of the phosphodiester bond dominates over peptide backbone fragmentation. As previously noted the substitution of the labile phosphate group blocks this preferential dissociation event. In consequence; the sequence information content is in general improved facilitating phosphorylation site determination. Representative MALDI-TOF/TOF data obtained from the α-S1 casein tryptic isolate at *m/z* 1770.3 which reflect this benefit are illustrated in Fig. 3. The resultant product ion spectrum displayed a nearly uninterrupted y ion series produced in high abundance. The location of the modification could be readily identified by the unique residue mass of 234 Da of the fragment ion y5 and hence serine in position 130 was recognized as the site of

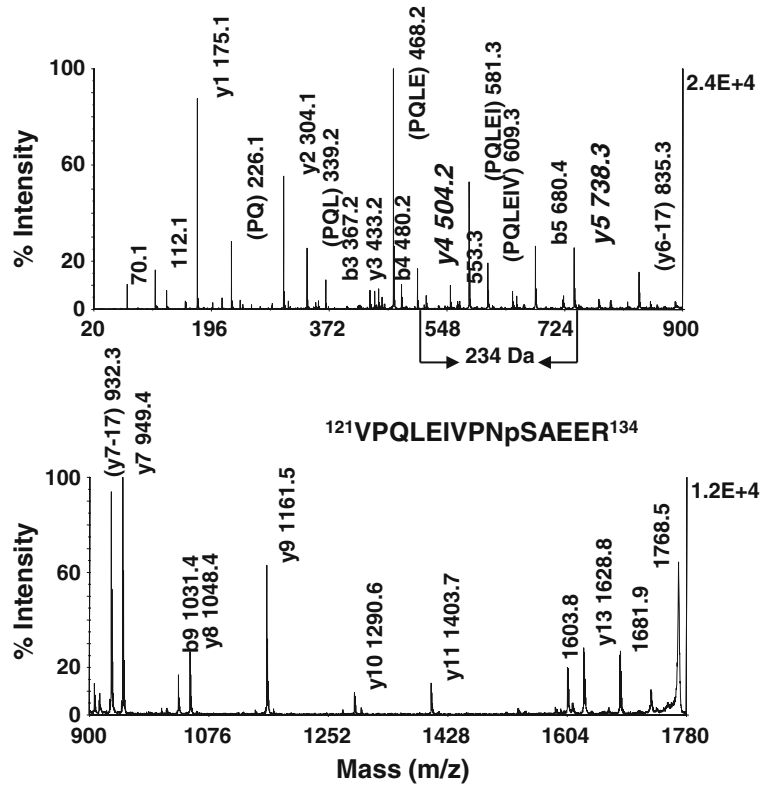


Fig. 3 Phosphorylation site determination by MALDI tandem MS. MALDI-TOF/TOF spectrum of thiol derivative at *m/z* 1770.3 enriched from α-S1/β-casein tryptic digest. The derivatization discriminates the site of phosphorylation as the unique residue mass of 234 Da contained in the product ion y5 *highlighted in italics* as is y4 which is contiguous to the modification. The presence of this ion pair enables unambiguous assignment of serine in position 130 as the site of phosphorylation. ~ 2 pmol of isolate were applied to the target

phosphorylation. Substituted phosphothreonine is recognized by its characteristic signature mass of 248 Da. The label remained stable under the condition of CID.

3.6 Diagnostic Tests

The tests are intended to monitor the efficiency of the individual reaction steps of the sample preparation method. In this way chemistry deficiency can be readily identified and remedies implemented to ensure that the final reaction products are formed to near completion.

1. Test #1. Performic acid oxidation/acetylation/BEMAD. Optimal formation of the intermediates and the final reaction product is desirable to ensure for method selectivity and to maximize the recovery of the Michael addition product. The data obtained from the application of the protocol described in **steps 1–3,** Subheading 3.3 using DAM1 phosphopeptide

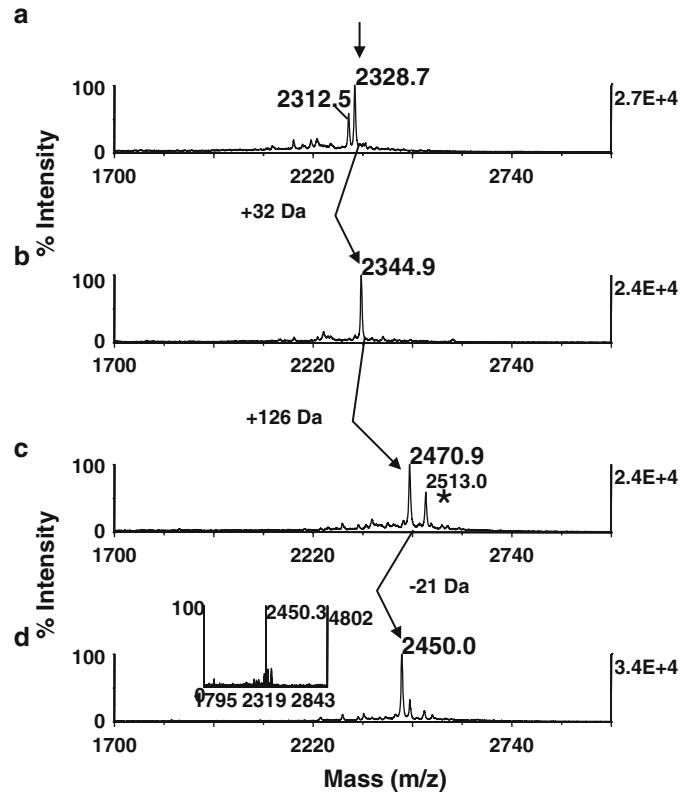


Fig. 4 Diagnostics test #1. MALDI MS spectra of (**a**) native DAM1 phosphopeptide SFVLNPTNIGMpSKSSQGHVTK, *m/z* 2312.5 (20 pm); (**b**) after oxidation; (**c**) after acetylation; (**d**) after BEMAD. *Cross arrows* denote mass shifts. *Arrow* and *asterisk* designate sulfoxide derivative and O-acylation product, respectively. ~2 pmol of peptide were applied to the target. *Inset* shows experiment at 300 fmol sample load using a μ-$_{C18}$ ZipTip from which the product was deposited in matrix onto the target

SFVLNPTNIGMpSKSSQGHVTK, *m/z* 2312.5 as test peptide are shown in Fig. 4a–d. Accordingly, the peptide was fully oxidized (*see* Fig. 4b), its primary amino groups were effectively acetylated (*see* Fig. 4c) and BEMAD resulted in nearly quantitative formation of the *N*-acetyl *S*-2-aminoethylcysteine adduct concomitant with complete O-acylation reversal (*see* Fig. 4d). Incompleteness of oxidation was observed only when the reagent was not allowed to be fully formed (*see* **Note 1**). Partial acetylation was strictly avoided by the immediate use of freshly prepared reagent (*see* **Note 2**). We have not encountered any complications during method optimization associated with BEMAD with respect to efficiency of derivatization or potential chemistry side reactions [11]. Inset shows the spectrum of the Michael addition product produced from 300 fmol of starting material. Consistent with previous results,

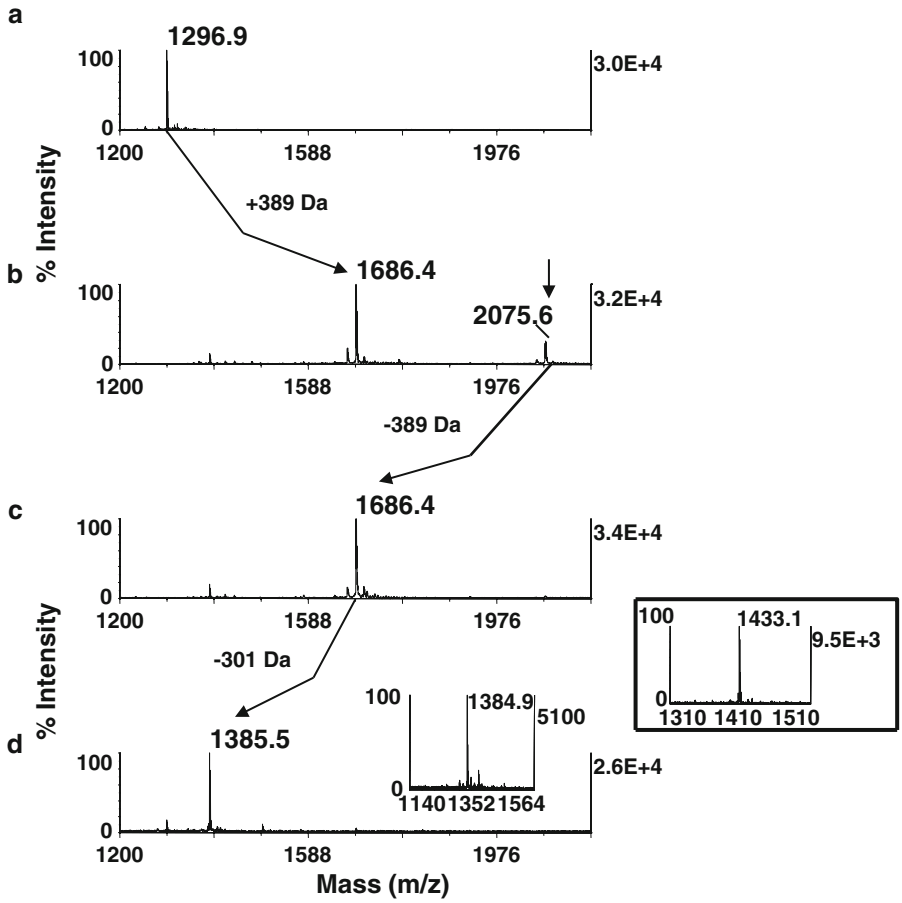


Fig. 5 Diagnostic test #2. MALDI MS spectra of (**a**) native angiotensin I DRVYIHPFHL, *m/z* 1296.9 (20 pm); (**b**) after biotinylation; (**c**) after O-deacylation; (**d**) after disulfide reduction. *Cross arrows* indicate mass shifts. *Arrow* denotes O-acylated peptide. *In-boxed area* shows sulfone derivative of thiolated angiotensin. ~2 pmol of peptide were applied to the target. *Inset* shows experiment at 500 fmol sample load using a μ-$_{C18}$ ZipTip from which the product was deposited in matrix onto the target

the data show that low-level amounts of phosphopeptide expected to occur in in vivo phosphorylated samples can be carried undiluted through the serial reaction scheme [11] (*see* **Note 16**).

2. Test #2. Another key feature of the sample preparation method is the N-thiolation sequence of serial reactions which imparts the affinity label onto the analyte. Results from the application of the N-thiolation procedure to the model peptide angiotensin I-DRVYIHPFHL, *m/z* 1296.5 are shown in Fig. 5a–d. Under the conditions described in **steps 4–6,** Subheading 3.3, the peptide was conjugated with Sulfo-NHS-SS-biotin to near completion (*see* Fig. 5b), its O-acylation product effectively hydrolyzed by hydroxylamine (*see* Fig. 5c), and the thiol-functionalized peptide analog produced in nearly quantitative

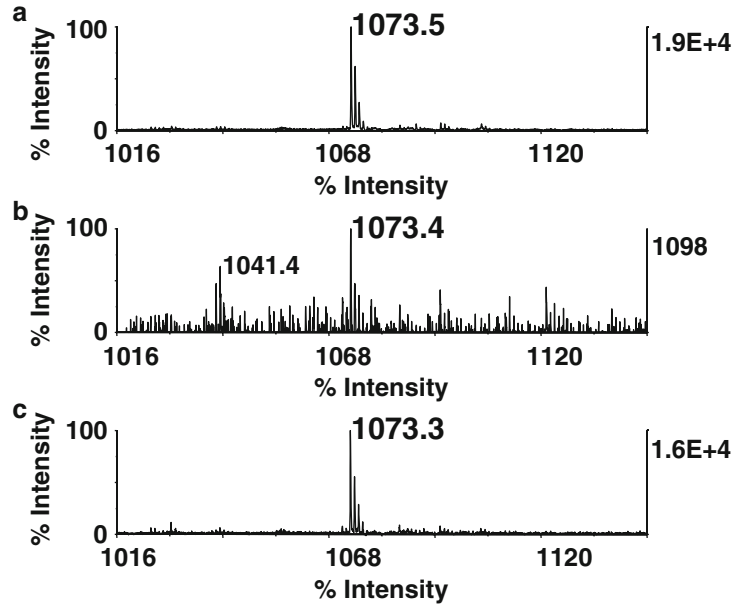


Fig. 6 Diagnostic test #3. Validation of analyte capture/release efficiency by covalent chromatography using the somatostatin fragment 3–10 (CKNFFWKT, *m/z* 1073.2) as test peptide. MALDI MS spectra of (**a**) starting material (5 pmol); (**b**) peptide recovered from the flow-through fraction; (**c**) peptide reductively released from the affinity support. ~1 pmol of peptide was analyzed. For MALDI MS, 80 laser shots acquired from eight different spot positions were summed for each spectrum

yield (*see* Fig. 5d). Inset shows the spectrum of the thiolated product produced from 500 fmol of starting material highlighting the advantage of solid phase reaction format to thiolate sub-picomole amounts of material. Failure to appreciate the recommendation with regard to the use of the NHS-ester reagent may lead to incomplete N-thiolation. As a result, the population of non-reacted Michael addition products would be subject to depletion from the affinity support resulting in diminished recovery of the isolates.

3. Test #3. The capability of the covalent chromatographic system to reversibly capture the analyte at high efficiency has been previously demonstrated with the cysteinyl peptide Somatostatin CKNFFWKT, *m/z* 1073.2 [11]. Results from such experiment are shown in Fig. 6a–c. Fractions collected before and after the reduction step were analyzed by MALDI MS along with the unprocessed sample. The MALDI MS spectra show that the signals produced from the starting material (*see* Fig. 6a) and from the reductively retrieved peptide (*see* Fig. 6c) were of comparable abundance. <10 % of the starting material was found in the initial flow-through fraction indicating

that a ~90 % coupling/release efficiency was achieved (*see* Fig. 6b). Variations between replicate experiments were <5 %. As noted, storage of the medium in the presence of bacteriostatics should be avoided as these agents impair the capture efficiency of the medium.

4 Notes

1. To avoid incomplete oxidation make sure that the mixture is left standing for at least 30 min prior to dilution to allow the reagent to be fully formed.
2. NHS-ester reagents in general have a half-life of 1 h at pH 8.0 and 25 °C. To ensure for completeness of reaction do not store the freshly prepared Sulfo-NHS acetate solution for more than 10 min prior to use. NHS-ester reagents are moisture sensitive. To prevent decomposition let the reagent equilibrate in the closed container from its low storage temperature (i.e., –20 °C) to room temperature. Obviously, do not prepare reagent in amine-containing buffers.
3. Grind barium hydroxide octahydrate crystals to a fine powder in a standard porcelain mortar. Transfer 31.54 mg of the chemical to a 1.5 mL microcentrifuge tube. Add 1 ml water and vortex vigorously for ~2 min to facilitate dissolution. Centrifuge for 1 min at 1811 × *g* to remove carbonate precipitates.
4. Considerations with regard to the proper use of Sulfo-NHS-SS-Biotin are as described in **Note 3**.
5. Do not add bacteriostatics such as sodium azide, merthiolate, or phenyl mercuric salts to the gel. These agents diminish the medium's binding capacity.
6. We used GelCode Blue Stain Reagent. We have thus far not employed other more sensitive non-covalent staining methods. We refer the reader to the report of Granvogl and coworkers in which the use of those methods is reviewed in ref. 17.
7. Standard protein in-gel digestion and in solution digestion protocols recommend alkylation after cystine reduction to promote the accessibility of the substrate to digestion. However, alkylated cysteinyl peptides are subject to BEMAD resulting in their co-isolation. This complication is addressed in our sample preparation method by post-digestion performic oxidation which renders cysteine inert toward BEMAD. Optionally, the solid phase oxidation step can be replaced by in-gel performic acid oxidation. The protein is then digested by the protease in the gel matrix. Gel extracts are bound to the reversed phase support and subsequently

subjected to acetylation (*see* Fig. 1). An effective gel-based procedure which we developed is found in the literature [11].

8. Monitor the progress of solvent evaporation. Do not dry the sample below the level indicated on the collection tube (~35 μL). Further solvent evaporation may cause substantial adsorptive peptide loss especially as seen with low-level samples up to 50 % or more of the starting solution [16].
9. The ZipTip de-wets in <2 min and should therefore be immediately processed for derivatization. Optionally, the ZipTip can be temporarily stored in 0.1 % TFA at 4 °C preferably for <60 min or at room temperature for <30 min. To this purpose place 60 μL 0.1 % TFA into 0.5 mL microcentrifuge tubes. Aspirate 10 μL of solvent onto the ZipTips. Leave the ZipTips immersed in solvent during storage.
10. Under the condition of concurrent BEMAD phosphothreonyl peptides are recalcitrant to derivatization as are phosphoseryl/threonyl peptides in which the phosphorylated residue is positioned adjacent to proline. The consecutive reaction mode has to be used to ensure efficient conversion of this class of phosphopeptides. During incubation air-induced carbonate precipitation typically occurs which has no impact on the efficiency of derivatization.
11. MALDI MS of the eluent furnishes the mass map of the starting material presented to covalent chromatography. Deviations from the anticipated result shown in Fig. 2a are typically due to incomplete digestion and/or incomplete derivatization. Do not refreeze the remainder of the enzyme solution for further use. Repeated thawing and freezing may impair enzyme activity. The presence of satellite peaks would indicate incompleteness of oxidation and/or acetylation. Remedies for these complications are discussed under Subheading 3.6.
12. Make sure that the affinity resin is properly washed to avoid cross contamination of the enriched fraction with non-phosphorylated peptides. Repeat the organic solvent wash if this problem persists.
13. To avoid adsorptive sample loss of the isolate do not dry the sample below the level marked on the collection tube (~50 μL) [16].
14. We recommend the use of ZipTipμ-$_{C18}$ pipette tips to recover phosphopeptide isolates enriched from low-level digests. With μ-tips the isolates can be deposited in 0.5 μL matrix directly on the MALDI-target. In this manner, femtomole mass detection can be readily achieved (*see* Figs. 2 and 5, insets).
15. In-solution tryptic digestion was performed in 40 μL of 25 mM ammonium bicarbonate/0.01 % OGS at an enzyme to substrate ratio of 1:100. After 18 h incubation at 37 °C, the

digest was acidified with 5 μL of 10 % TFA prior to solid phase immobilization.

16. Any other medium-size phosphoseryl peptide containing methionine may be used as test peptide.

Acknowledgements

This work has been funded by NIH grants R33CA101150 and P20-DA026149 to R.H.A. The authors thank the Albert Einstein College of Medicine and the MD Anderson Cancer Center for generous support, Dr. Richard Stanley for helpful discussions and Ms. Junko Hihara for editorial assistance.

References

1. Helmbrecht K, Zeise F, Rensing L (2000) Chaperones in cell cycle regulation and mitotic signal transduction: a review. Cell Prolif 33:341–356
2. Posewitz MC, Tempst P (1999) Immobilized gallium (III) affinity chromatography of phosphopeptides. Anal Chem 71:2883–2892
3. Klemm ZC, Otto S, Wolf S et al (2006) Evaluation of the titanium dioxide approach for MS analysis of phosphopeptides. J Mass Spectrom 41:1623–1632
4. McLachlin TD, Chait BT (2003) Improved β-elimination-based affinity purification strategy for enrichment of phosphopeptides. Anal Chem 75:6826–6836
5. Brocklehurst K, Carlsson J, Marek P et al (1973) Covalent chromatography. Preparation of fully active papain from dried papaya latex. Biochem J 133:573–584
6. Amoresano A, Marino G, Cirulli C et al (2004) Mapping phosphorylation sites: a new strategy based on the use of isotopically labeled DTT and mass spectrometry. Eur J Mass Spectrom 10:401–412
7. Vosseler K, Hansen KC, Chalkley RJ et al (2005) Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. Proteomics 5:388–389
8. Xiao J, Nieves E, Angeletti RH et al (2006) Rat organic anion transporting protein 1A1 (Oatp 1a1): purification and phosphopeptide assignment. Biochemistry 45:3357–3369
9. Salih E, Siqueria WL, Helmerhorst EJ et al (2010) Large-scale phosphoproteome of human whole saliva using disulfide-thiol-interchange covalent chromatography and mass spectrometry. Anal Biochem 407:19–33
10. Wells L, Vosseller K, Cole RN (2002) Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. Mol Cell Proteomics 1:791–804
11. Nika H, Nieves E, Hawke DH et al (2013) Phosphopeptide enrichment by covalent chromatography after derivatization of digests immobilized on reversed phase supports. J Biomol Tech 24:54–177
12. Sweiger R, Linial M (2010) Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data. Biol Direct 5(6):1–17
13. Bodenmiller B, Mueller LN, Pedrioli PG et al (2007) An integrated chemical, mass spectrometric and computational strategy for phosphoproteomics: application to Drosophila melanogaster Kc167 cells. Mol Biosyst 3:275–286
14. Johnson M, Carpenter TS (2005) The use of solid-phase supports for derivatization in chromatography and spectroscopy. Appl Spectrosc Rev 40:391–412
15. Zhang Q, Qian W-S, Knyushko TV et al (2007) A method for selective enrichment and analysis of nitrotyrosine-containing peptides in complex proteome samples. J Proteome Res 6:2257–2268
16. Speicher DT, Kolbas O, Harper D et al (2000) Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. J Biomol Tech 11:74–86
17. Granvogl B, Ploscher M, Eichacker LA (2007) Sample preparation by in-gel digestion for mass spectrometry-based proteomics. Anal Bioanal Chem 389:991–1002

Part II

Labeling Techniques for Quantitative Phosphoproteomics

Chapter 4

Peptide Labeling Using Isobaric Tagging Reagents for Quantitative Phosphoproteomics

Lei Cheng, Trairak Pisitkun, Mark A. Knepper, and Jason D. Hoffert

Abstract

Isobaric tagging reagents have become an invaluable tool for multiplexed quantitative proteomic analysis. These reagents can label multiple, distinct peptide samples from virtually any source material (e.g., tissue, cell line, purified proteins), allowing users the opportunity to assess changes in peptide abundances across many different time points or experimental conditions. Here, we describe the application of isobaric peptide labeling, specifically *8plex* isobaric tags for relative and absolute quantitation (8plex iTRAQ), for quantitative phosphoproteomic analysis of cultured cells or tissue suspensions. For this particular protocol, labeled samples are pooled, fractionated by strong cation exchange chromatography, enriched for phosphopeptides, and analyzed by tandem mass spectrometry (LC-MS/MS) for both peptide identification and quantitation.

Key words IMAC, Isobaric tags, Isotopic labeling, iTRAQ, LC-MS/MS, Mass spectrometry, Multiplexing, Phosphopeptide, Phosphoproteomics, Reporter ion, TMT

1 Introduction

Reversible protein phosphorylation is a key post-translational modification responsible for various cellular regulatory mechanisms. Protein phosphorylation studies are challenging since phosphorylated proteins are often low in abundance and of low stoichiometry. Moreover, phosphorylated peptides from a mixture often exhibit low ionization efficiencies during LC-MS/MS analysis due to ion suppression effects. Thus, careful sample preparation, adequate sample amount, and efficient phosphopeptide enrichment steps are basic requirements for any successful phosphoproteomic analysis.

Phosphopeptide enrichment methods are widely adapted to the "bottom up" proteomics approach which is characterized by proteolytic digestion of proteins into peptide fragments prior to analysis by mass spectrometry. Immobilized metal affinity chromatography (IMAC) is based on the affinity of the negatively charged

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_4,

phosphate groups on phosphopeptides for a positively charged metal ion column matrix, and it remains the most widely used method for affinity enrichment [1–3]. However, metal oxides, especially titanium dioxide (TiO_2), are common alternatives to IMAC and often can isolate unique subsets of phosphopeptides not enriched by other methods [4, 5].

Novel MS acquisition techniques have also spurred growth in the field of phosphoproteomics. Techniques such as neutral loss scanning, precursor ion scanning, and multi-stage activation (MSA) have been successfully applied to the routine identification of protein phosphorylation from complex biological samples [6–8]. New fragmentation methods including HCD, ECD, and ETD have also been utilized for protein phosphorylation analysis, which has allowed better fragmentation of the phosphorylated peptides, improved assignment of phosphorylation sites, and increased the sensitivity of MS-based protein phosphorylation analysis [9–11].

One of the breakthroughs in the field of proteomics has been the development of a vast array of quantitative methods. These include various label-free methods, stable isotope labeling, and targeted quantification techniques. All methods are applicable to phosphoproteomics, and quantitative phosphoproteomics has become an important method for measuring changes in protein phosphorylation on a global scale. Stable isotope labeling approaches generally produce more reliable quantification results compared to label-free quantification. Stable isotope labeling strategies include stable isotope labeling by amino acids in cell culture (SILAC), dimethyl labeling, and the use of isobaric tagging reagents such as isobaric tags for relative and absolute quantitation (iTRAQ) and the tandem mass tagging (TMT) approach. Although all three methods have their strengths and weaknesses, a recent study indicates that all three can reach a similar level of sensitivity based on the number of identified proteins using a classical (MS2-based) shotgun approach [12]. SILAC and dimethyl labeling strategies quantify peptides at the MS1 level. The more differential labels are used, the more complex the MS1 spectra will be. Thus, normally only two or three differential labels are used. The isobaric tagging strategy, on the other hand, quantifies peptides at the MS2 level. Differentially labeled peptides will have the same m/z (at the MS1 level) and will be selected for MS2 analysis at the same time. Therefore isobaric labeling can allow quantitative comparison of up to ten different peptide samples, e.g.,using the commercially available TMT 10plex kit. It is worth noting reports of the use of hyperplexing (i.e., 18-plex), as well as a more recent 54-plex technique, which have greatly enhanced the capacity for sample multiplexing with isobaric reagents [13, 14]. In addition to the advantage provided by multiplexing, isobaric tagging approaches are relatively easy to perform. Furthermore, these approaches can be

adapted to label virtually any sample type (e.g., cell line, tissue, or purified proteins).

The isobaric tagging approach is based on the covalent labeling of the N-terminus and side-chain primary amines of peptides with tags of varying masses through NHS-ester chemistry, followed by MS analysis [15, 16]. The structure of each reagent consists of three distinct regions: (1) a cleavable reporter group of a specific mass for peptide quantitation (113, 114, 115, 116, 117, 118, 119, and 121 Da in the case of 8plex iTRAQ), (2) a mass normalizer or "balancer" region that makes each tag isobaric, and (3) an amine reactive group that will covalently attach the tag to the peptide (*see* Fig. 1). Relative quantification of a peptide is based on different reporter ions generated in the low mass area of its MS2 spectra (*see* Fig. 2). Due to the small size of the reporter ions, iTRAQ is compatible only with wider mass range instruments such as triple quadrupole and the Orbitrap generation of mass spectrometers, not with traditional ion traps. The signals of these reporter ions normally do not interfere with b and y ions used for peptide identification. Peptide samples to be labeled with isobaric tagging reagents should be free of the following: thiols, high concentrations of detergents or denaturants, and chemicals/buffers with primary amines other than the analyte of interest. Primary amines can react with the isobaric tagging reagents resulting in insufficient labeling of sample peptides. Equal amounts of labeled samples are then pooled, fractionated and enriched for phosphopeptides, and followed by LC-MS/MS analysis. The same peptide from differentially labeled samples will still possess the same mass, i.e., the original mass plus the mass of the isobaric tag less one proton due to conjugation (+304 Da in the case of 8plex iTRAQ). Thus, the isobaric tagging approach does not lead to more complex MS1 spectra as the differentially labeled peptides co-elute from the HPLC prior to MS analysis. During LC-MS analysis, these peptides are co-isolated for MS/MS fragmentation, where they generate the same b and y ion series for peptide identification while the relative quantification information is retained in the ratios of the reporter ion series. The fact that isobaric tagging reagents allow multiplexing is advantageous for research projects involving a time course design, e.g., monitoring changes in protein expression or changes in the level of various post-translational modifications following hormone stimulation across different time points or biological conditions. The labeling step for the isobaric tagging approach is performed after protein digestion, thus any variability in sample handling prior to sample pooling will increase the quantification biases. A normalization procedure can be adapted to correct for these quantification errors (*see protocol below*). Once the labeled peptides are pooled, further experimental biases will be minimized. For phosphoproteomics workflows in particular, the fractionation step as well as the phosphopeptide

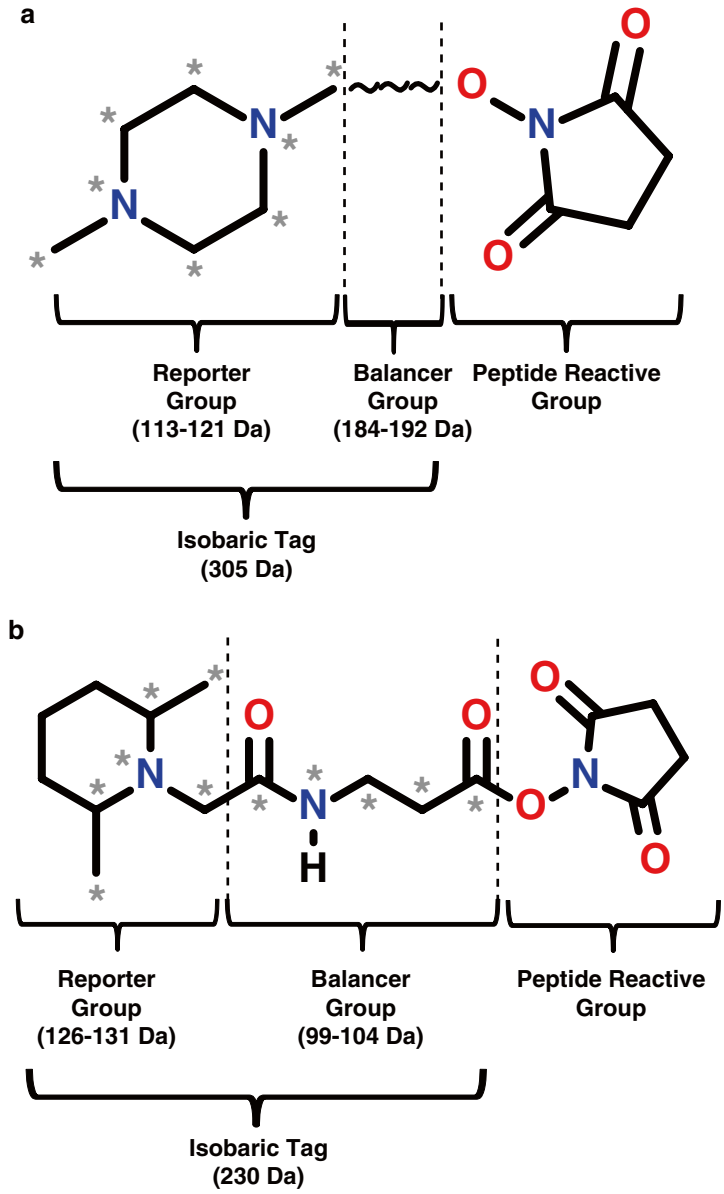


Fig. 1 Chemical structures of isobaric tagging reagents. The general structure for both 8plex iTRAQ (**a**) and 6plex TMT (**b**) tags consists of an MS-cleavable reporter group, a balancer group of variable sizes to make the tag isobaric, and a peptide reactive group for labeling. *Asterisks* indicate positions of ^{13}C and ^{15}N heavy isotope substitutions which are used to generate reporter ions of various sizes. *Vertical dashed lines* indicate bonds that break during labeling (*right-hand lines*) and bonds that break during MS fragmentation (*left-hand lines*). Note: the structure of the 8plex iTRAQ balancer group is not yet published

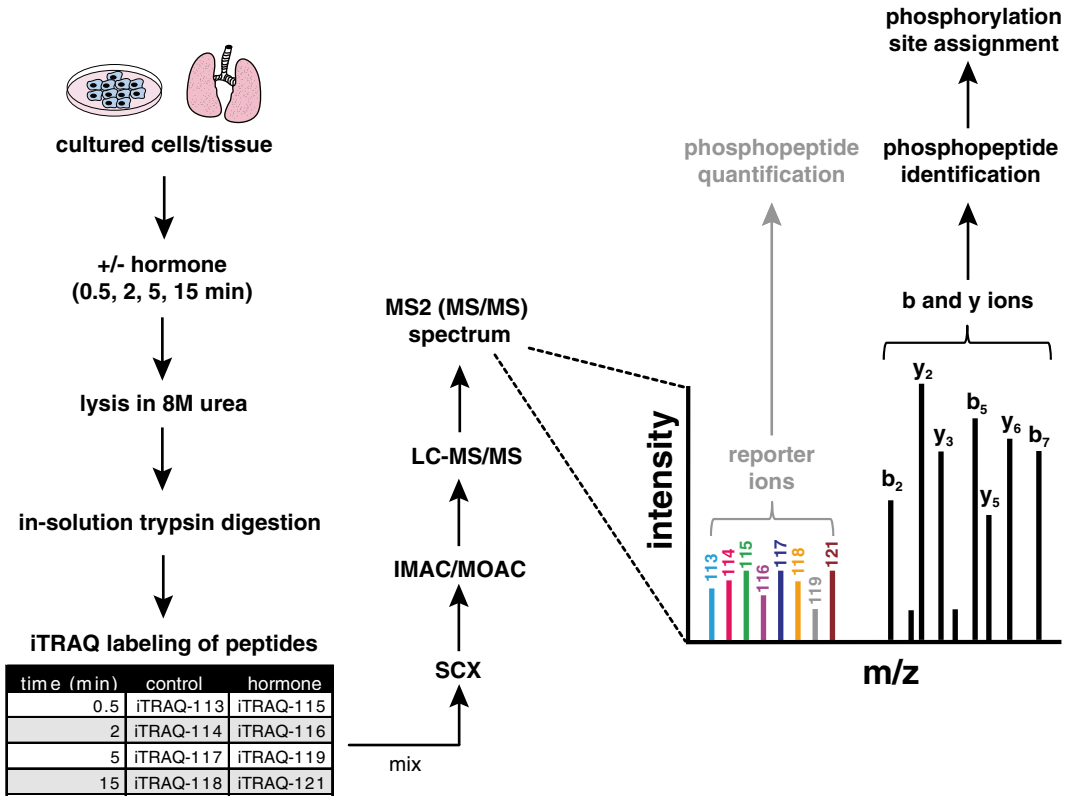


Fig. 2 Experimental workflow for iTRAQ-based quantitative phosphoproteomic analysis. Cultured cells/tissue suspensions are treated with or without hormone for the indicated times followed by lysis in 8 M urea. Protein lysates are then digested with trypsin, desalted, and labeled with 8plex iTRAQ reagents. Strong cation exchange chromatography (SCX) stratifies the sample into 20 fractions followed by either immobilized metal affinity chromatography (IMAC) or metal oxide affinity chromatography (MOAC), which will enrich each fraction for phosphopeptides. Phosphopeptides are analyzed by tandem mass spectrometry (LC-MS/MS) in which the fragmentation is performed by higher energy collision induced dissociation (HCD), and the mass-to-charge ratio (*m*/*z*) and intensity of corresponding peptide ions are measured by an orbitrap-based mass spectrometer. In the MS2 spectrum, the pattern of b and y ions allows for phosphopeptide identification through database searching (*black peaks*), while the intensities of the iTRAQ reporter ions allow for relative quantification of phosphopeptide abundances across the eight different experimental conditions (*colored peaks*)

enrichment step should not introduce additional quantification errors since the samples should have already been pooled before these steps. There are currently two types of iTRAQ reagents available: 4plex and 8plex. With the 4plex reagent, up to four different biological conditions can be investigated at the same time; and with the 8plex reagent, up to eight. The 4plex and 8plex reagents have different structures in the balancer group region; however, they show only slight differences in sensitivity. There was an initial report that showed that 4plex kits may generate higher numbers of protein identifications compared to 8-plex kits [17]. However, it

was later shown that 8plex iTRAQ provides more consistent quantification ratios compared to 4plex, and provides comparable total identifications while allowing more experimental conditions to be investigated in a large scale proteomics study [18].

One common problem encountered during LC-MS/MS analysis of iTRAQ or TMT labeled complex samples is the co-isolation of contaminant ions with similar m/z values and elution times. This means that for a given peptide, the reporter ion ratios in its corresponding MS2 spectrum do not reflect the true quantification ratios for that peptide, but instead reflect the sum of all reporter ion intensities produced by that peptide and from all other contaminating peptides co-isolated with the peptide of interest. This phenomenon is often referred to as "isolation interference" or "ratio compression," as it tends to compress peptide quantitation ratios toward unity (i.e., 1). The problem can be partially alleviated by performing fractionation at the peptide level using techniques such as SCX or HILIC chromatography. Fractionation reduces the complexity of the original sample and is usually based on an alternative peptide separation strategy other than C18 (normally the method of choice for HPLC separation coupled to MS analysis). At the data analysis level, software such as Proteome Discoverer (Thermo Scientific) can calculate isolation interference scores based on the unassigned peaks and their intensities presented in MS2 spectra. Using an appropriate isolation interference score cutoff, users can filter large-scale iTRAQ quantification data with more reliable results. At the MS acquisition level, two MS techniques have been adopted for overcoming the ratio compression problem: gas phase fractionation [19] and MS3 acquisition [20]. Gas phase fractionation uses the proton-transfer ion-ion reactions (PTR) to reduce the precursor ion charge state and gets rid of contaminating ions with different charge states. MS3 acquisition provides an additional isolation and fragmentation event that helps minimize the interference problem. However, it was noted that the MS3 method suffered from reduced sensitivity. To overcome issues with sensitivity, a relatively recent approach was developed called Synchronous Precursor Selection (MultiNotch) MS3 which allows isolation of multiple MS2 product ions simultaneously, helping to increase the intensity of reporter ions in MS3 spectra and improving sensitivity, precision, and accuracy in MS quantification [21].

In this chapter we will introduce a standard workflow for 8plex iTRAQ labeling of peptides isolated from mammalian cells or tissue suspensions for multiplexed quantitative phosphoproteomic analysis. A similar workflow was recently used to successfully probe the phosphorylation dynamics of the vasopressin V2 receptor signaling pathway in mammalian kidney [22].

2 Materials

Note: All reagents including water, acetonitrile, and isopropanol should be HPLC-grade or higher.

2.1 Preparation of Cell Lysates

1. Cells.
2. Hormone for stimulation (For example: vasopressin).
3. Cell Lysis Buffer, 8 M urea, 50 mM Tris–HCl, 75 mM NaCl, 1× Halt Protease and Phosphatase Inhibitor Cocktail.
4. Benchtop Centrifuge.
5. Probe sonicator (Misonix 3000 or equivalent).
6. Reagents for protein assay (e.g., BCA assay).
7. 1.5 ml microcentrifuge tubes.

2.2 In-Solution Protease Digestion

1. 50 mM Ammonium Bicarbonate (AmBic) Buffer, 0.2 g ammonium bicarbonate in 50 ml HPLC-grade water.
2. 250 mM Dithiothreitol (DTT) stock, 7.7 mg DTT in 200 µl AmBic.
3. 250 mM Iodoacetamide stock, 9.3 mg iodoacetamide in 200 µl AmBic.
4. 1 µg/µl Trypsin stock, 100 µg Trypsin Gold in 100 µl of 50 mM acetic acid. Keep on ice until ready to use, then freeze the unused portion at –20 °C.
5. 100 % Formic Acid.
6. Benchtop Centrifuge.
7. pH meter or pH paper.
8. Waters Oasis HLB 1 cc Desalting Cartridges (WAT094225 or equivalent).
9. 100 % Acetonitrile (ACN).
10. Water (LC/MS grade).
11. Savant SC100 SpeedVac with RT490 Refrigerated Condensation Trap.

2.3 iTRAQ Labeling

1. iTRAQ 8plex Multi-plex Kit (AB SCIEX): 5× 1-U vials of each iTRAQ 8plex reagent (i.e., 113, 114, 115, 116, 117, 118, 119, and 121), Dissolution Buffer pH 8.5 (0.5 M triethylammonium bicarbonate, TEAB), and isopropanol. *Important note: The denaturant, reducing reagent, and cysteine-blocking reagent vials provided with this kit are not used in this protocol.*
2. 100 % Formic Acid.
3. Savant SC100 SpeedVac with RT490 Refrigerated Condensation Trap.

4. pH meter or pH paper.
5. 15 ml conical tubes.

2.4 Sample Fractionation by Strong Cation Exchange Chromatography

1. PolySulfoethyl A SCX column (4.6 mm ID × 20 cm length, 5-μm particle size, 300-Å pore size; PolyLC).
2. SCX Buffer A, 5 mM KH_2PO_4/25 % ACN, pH 2.67. *Dissolve 0.68 g KH2PO4 in 747 ml LC-MS/MS grade water. Monitoring with a pH meter and with constant mixing, add ~2.5–3 ml of 1 N HCl to bring pH to 2.67. Add 250 ml 100 % ACN and mix.*
3. SCX Buffer B, 5 mM KH_2PO_4/500 mM KCl/25 %ACN, pH 2.67. *Dissolve 0.68 g KH2PO4 and 37.29 g KCl in 747 ml LC-MS/MS grade water. Monitoring with a pH meter and with constant mixing, add ~2.5–3 ml of 1 N HCl to bring pH to 2.67. Add 250 ml 100 % ACN and mix.*
4. HPLC system (Agilent HP1100 System or equivalent).
5. Waters Oasis HLB cartridge.

2.5 Phosphopeptide Enrichment

1. Pierce Fe-NTA Phosphopeptide Enrichment Kit (Pierce/Thermo).
2. Pierce Graphite Spin Columns (Pierce/Thermo).

2.6 LC-MS/MS Analysis

1. Eksigent Nanoflow LC system connected to an LTQ Orbitrap Velos mass spectrometer or an equivalent LC-MS/MS system.
2. MS Buffer A: 0.1 % formic acid in water.
3. MS Buffer B: 0.1 % formic acid in acetonitrile.

2.7 Phosphopeptide Identification (Protein Database Searching)

1. Proteome Discoverer Software (or equivalent)

3 Methods

3.1 Preparation of Cell Lysates

1. Incubate cell line/tissue suspensions with hormone/reagent of choice for the appropriate amounts of time. The amount of protein for each sample should be at least 100 μg (optimally 500 μg) for each desired experimental condition. A typical 8 plex iTRAQ time course experimental design is provided in Fig. 2 (*see* **Note 1**).
2. Following incubation, briefly spin samples at 10,000 × *g* for 30 s to pellet the cells and remove the supernatant.
3. Resuspend cell pellets in 150 μl of Cell Lysis Buffer in a 1.5 ml microcentrifuge tube.

4. Place samples in a small container of wet ice. Sonicate immediately using a Misonix probe sonicator or equivalent for 1 min, setting 1, with 0.5 s bursts.
5. Spin at >10,000 × *g* for 10 min in a benchtop centrifuge to pellet cellular debris. Transfer the supernatants to new microcentrifuge tubes.
6. Perform a protein assay (e.g., BCA assay). The samples should contain at least 100 μg (optimally 500 μg) of protein and the concentration should be approximately 4 μg/μl (*see* **Note 2**).

3.2 In-Solution Protease Digestion

1. Reduce the samples by adding DTT to a final concentration of 10 mM. Incubate 1 h at 37 °C.
2. Alkylate the samples by adding iodoacetamide to a final concentration of 40 mM. Incubate 1 h, no longer. (Protect sample from light.)
3. Quench the excess iodoacetamide by adding another 40 mM DTT. Incubate for at least 15 min at room temperature.
4. Dilute the samples to <1 M urea with 50 mM AmBic.
5. Add trypsin at a trypsin-to-protein ratio of 1:20 to 1:100 (weight: weight). Ideally, the final trypsin concentration in the sample should be ≥12 ng/μl. Incubate at 37 °C for 16 h.
6. Terminate the reaction by adding 100 % formic acid to a final concentration of 0.5 %.
7. Spin the samples at ≥16,000 × *g* for 20 min at 4 °C in a benchtop centrifuge to pellet any insoluble material. Transfer the supernatants to fresh tubes. Check that the pH is <4.0.
8. Desalt the samples using a Waters Oasis HLB cartridge (*see* **Note 3**).
   (a) Condition the cartridge with 1 ml of 100 % ACN.
   (b) Equilibrate with 1 ml of 0.1 % formic acid.
   (c) Slowly apply the peptide sample to the cartridge (1 drop every 3 s).
   (d) Wash the cartridge three times with 1 ml of 0.1 % formic acid.
   (e) Elute the desalted peptides slowly using 1 ml of 0.1 % formic acid/50 % ACN.
   (f) Vacuum-concentrate the samples down to <10 μl using a SpeedVac.

At any step in the protocol that includes vacuum concentration of peptides, samples can be stored at ≤ −20 °C.

Vacuum concentration using a SpeedVac is often a slow process, especially for larger volumes or less volatile liquids. For convenience, samples can be safely left overnight in the SpeedVac without compromising the integrity of the peptide sample.

3.3 iTRAQ Labeling

(Note: The following protocol is for labeling 500 μg of peptide per iTRAQ channel. At least 100 μg of peptide per iTRAQ channel should be used. Please scale the amount of each reagent accordingly.)

1. Bring iTRAQ reagent vials, Dissolution Buffer, and isopropanol to room temperature (*see* **Note 4**).
2. Preparation of iTRAQ reagents.
   (a) Briefly spin iTRAQ reagent vials to bring the liquid to the bottom of tube.
   (b) Add 70 μl of isopropanol to each vial. Vortex and spin.
   (c) Combine the contents of the five duplicate iTRAQ reagent vials into a single vial for each reagent. Each iTRAQ reagent vial should now contain approximately 350–370 μl of reagent.
   (d) Vortex the tubes and spin again.
3. Resuspend the peptide samples in 150 μl of iTRAQ Dissolution Buffer.
4. Add the total contents of each iTRAQ reagent vial to each sample according to your particular experimental design (An example is provided in Fig. 2). Vortex briefly to mix (*see* **Note 5**).
5. Incubate for 2 h at room temperature.
6. Quench the reaction by adding formic acid to a final concentration of 0.5 %. Samples can be stored at −80 °C if necessary before proceeding with the rest of the protocol.
7. Vacuum-concentrate the samples to <50 μl to remove the majority of isopropanol. Important: Avoid letting samples dry completely or they will be difficult to resuspend during the next step.
8. Resuspend each sample in 500 μl of 0.5 % formic acid.
9. Combine all 8 iTRAQ-labeled samples into a single 15 ml conical tube. Check that the pH is <4.0.
10. Divide the sample equally across four desalting cartridges. Desalt the sample as in **step 8**, Subheading 3.2. Vacuum-concentrate the sample to a volume <10 μl.

3.4 Sample Fractionation by Strong Cation Exchange Chromatography

1. Resuspend the sample in 300 μl of SCX Buffer A. Check that the pH is 2.6–3.0.
2. Load the sample onto a conditioned PolySulfoethyl A SCX column attached to an HPLC system (Agilent HP1100 System or equivalent).
3. Run at a flow rate of 1 ml/min using the following gradient: 100 % buffer A and 0 % buffer B for 2 min; 0–14 % buffer B for 33 min; 14–100 % buffer B for 1 min; 100 % buffer B held for 4 min.

4. Collect fractions every 1.5 min. Based on the chromatographic profile at 214 nm, pool the samples down to 20 fractions (*see* **Note 6**).
5. Vacuum-concentrate the samples to a volume <10 µl. Resuspend samples in 0.1 % formic acid (*see* **Note 7**). Check that the pH is <4.0.
6. Desalt each fraction using a Waters Oasis HLB cartridge and reduce volume to <10 µl by vacuum-concentration (*see* **step 8**, Subheading 3.2).

3.5 Phosphopeptide Enrichment

1. Process all 20 SCX fractions by immobilized metal affinity chromatography (IMAC) or metal oxide affinity chromatography (MOAC) to enrich for phosphopeptides (*see* **Note 8**).
2. For IMAC, resuspend the labeled peptide samples in 200 µl of Binding Buffer (Pierce Fe-NTA Phosphopeptide Enrichment Kit).
3. Add sample to a Fe-NTA spin column and incubate for 20 min at room temperature with end-over-end rotation. Centrifuge the column at 1000 × *g* for 1 min. Discard the flow-through. Transfer column to a new tube.
4. Add 100 µl of Wash Buffer A to the spin column and gently mix the contents by tapping the side of the column. Do not pipette up and down.
5. Centrifuge the column at 1000 × *g* for 1 min. Discard the flow-through.
6. Repeat **steps 4** and **5** once.
7. Add 100 µl of Wash Buffer B to the spin column and gently mix the contents as before.
8. Centrifuge the column at 1000 × *g* for 1 min. Discard the flow-through.
9. Repeat **steps 7** and **8** once.
10. Add 100 µl of ultrapure water to the column and gently mix. Centrifuge the column at 1000 × *g* for 1 min. Discard the flow-through.
11. Transfer the column to a new collection tube and add 50 µl of Elution Buffer directly to the resin. Incubate for 5 min at room temperature.
12. Centrifuge the column at 1000 × *g* for 1 min. Retain eluate for analysis.
13. Repeat **steps 11** and **12**, Subheading 3.5, two additional times and pool the elution fractions.
14. Acidify the pooled elution by adding 200 µl of 2.5 % TFA.
15. Desalt samples using Pierce Graphite Spin Columns prior to analysis by mass spectrometry.

3.6 LC-MS/MS Analysis

1. Resuspend the desalted, phosphopeptide-enriched samples in 20 µl of 0.1 % formic acid.
2. Inject 10 µl of each sample onto an Eksigent Nanoflow LC system connected to an LTQ Orbitrap Velos mass spectrometer or an equivalent LC-MS/MS system (*see* **Note 9**). Save the other half of each sample for a subsequent LC-MS/MS run.
3. The following MS instrument parameters should be used: peptides ionized via a nano-spray ion source; MS run time of 65 min; spectra recorded in data-dependent acquisition mode with the dynamic exclusion option enabled; each survey MS scan followed by Higher Energy Collision Induced Dissociation (HCD) fragmentation of the top six most abundant precursor ions; both survey MS as well as MS2 scans acquired by the Orbitrap mass analyzer with a resolution of 30,000 and 7500 at m/z of 400 for MS and MS2 scans, respectively. For more effective fragmentation of iTRAQ-labeled peptides, set the normalized collision energy to 45 % (i.e., 10–15 % higher than for native peptides) or use a stepped normalized collisional energy scheme during HCD [23]. To minimize isolation interference, the precursor isolation window should be set to as narrow a width as possible (given that the sensitivity is not compromised). We recommend an isolation window of 3 m/z (i.e., ±1.5 m/z).

3.7 Phosphopeptide Identification (Protein Database Searching)

1. Search MS2 spectra (RAW files) using Proteome Discoverer Software running the Sequest search algorithm on a concatenated database containing both forward and reversed complement sequences from the latest version of the NCBI Refseq Protein Database from the appropriate species. Append a list of common contaminating proteins (e.g., porcine trypsin and human keratin) (http://www.thegpm.org/crap/) (*see* **Note 10**).
2. The following MS search parameters are recommended: precursor ion tolerance set to 25 ppm; fragment ion tolerance set to 0.05 Da; three missed trypsin cleavages; static modifications are carbamidomethylation of cysteine (+57.021 Da) and iTRAQ 8plex modification of lysine and peptide N-termini (+304.205 Da); variable modifications are oxidation of methionine (+15.995 Da), phosphorylation of serine, threonine, and tyrosine (+79.966 Da), and iTRAQ 8plex modification of tyrosine (+304.205 Da); target-decoy filter set to a 1 % false discovery rate (FDR) at the peptide level; known contaminant ions should be excluded. In addition, each batch of iTRAQ reagents contains trace levels of isotopic impurities. Thus, users should also set the *isotope correction factors* based on the values provided in the certificate of analysis that comes with each iTRAQ kit.

3. Phosphorylation sites should be assigned using a phosphorylation site assignment algorithm such as PhosphoRS (provided with Proteome Discoverer Software), PhosSA [24], or Ascore [25] (*see* **Note 11**).
4. Phosphopeptides that match to more than one protein isoform should be identified using programs such as MassSieve [26] and ProMatch [27]. Although it is not necessary to eliminate these peptide IDs from further analysis, iTRAQ quantification values obtained from these "ambiguous" peptides may reflect average peptide abundances from multiple protein isoforms that may be present in the sample.

3.8 Phosphopeptide Quantification

1. MS2 iTRAQ reporter ion intensities for phosphopeptides that possess the same linear amino acid sequence as well as the same site(s) of modification (including all types of modifications, not just phosphorylation) should be summed for each individual iTRAQ channel (*see* **Note 12**).
2. The desired relative abundance ratios are then calculated for each phosphopeptide (see the experimental design in Fig. 2). *For an arbitrary peptide X*:

$$\mathrm{iTRAQ-115\ peptide\ X}\left[\mathrm{hormone, 0.5min}\right] = 1000$$
$$\mathrm{iTRAQ-113\ peptide\ X}\left[\mathrm{control, 0.5min}\right] = 500$$

then:

$$\frac{\mathrm{iTRAQ-115\ peptide\ X}}{\mathrm{iTRAQ-113\ peptide\ X}} = 2$$

3. These values are then normalized using a global correction factor based on the ratio of the summed reporter ion intensities of all peptides in each corresponding iTRAQ channel. *If the summation of all reporter ion intensities for all peptides in each channel are*:

$$\mathrm{iTRAQ-115\ all\ peptides}\left[\mathrm{hormone, 0.5min}\right] = 4{,}000{,}000$$
$$\mathrm{iTRAQ-113\ all\ peptides}\left[\mathrm{control, 0.5min}\right] = 3{,}900{,}000$$

then the normalization factor is:

$$\frac{\mathrm{iTRAQ-115\ all\ peptides}}{\mathrm{iTRAQ-113\ all\ peptides}} = 1.03$$

and the normalized abundance ratio for peptide X is:

$$\frac{\mathrm{iTRAQ-115\ peptide\ Xnorm}}{\mathrm{iTRAQ-113\ peptide\ Xnorm}} = 2\left(\frac{1}{1.03}\right) = 1.95$$

4. The final step is to take the $\log_2$ of this normalized ratio.

$$\log_2(1.95) = 0.963$$

5. The $\log_2$ normalized ratio is then used to calculate the mean and standard deviation of the relative abundance of each peptide among all biological replicates (*see* **Note 13**).
6. A one-sample *t*-test can be used to calculate a *p*-value for each peptide. Specifically, all $\log_2$ normalized ratios for a given peptide are compared to a hypothetical mean of 0 [i.e., $\log_2(1) = 0$ is equivalent to a fold change of 1, or no change].
7. To correct for the higher number of false positive hits produced by multiple testing (i.e., thousands of peptides are routinely analyzed in a single data set), we recommend the use of a multiple testing correction method. The Benjamini and Hochberg (BH) False Discovery Rate [28] is relatively easy to calculate and represents an acceptable tradeoff between sensitivity and specificity (*see* **Note 14**). To calculate:
   (a) Rank the *p*-value of each peptide from smallest to largest. The smallest *p*-value has a rank of $r = 1$, the next has a rank of $r = 2$, etc.
   (b) Compare each peptide's *p*-value to $(r/n)\,Q$, where n is the total number of peptides and Q is the chosen FDR (usually 0.05 or less).
   (c) A *p*-value is considered significant (i.e., passed the FDR filter) if $p < (r/n)Q$.
   (d) For a list of ten peptides that will be filtered for a FDR (Q) value of 0.05 or 5 %, see the example below. In this case, only the top three peptides will pass the BH 5 % filter [$p < (r/n)Q$].

Peptide	Rank (r)	p-value	$(r/n)Q$
Peptide 1	1	0.001	0.005
Peptide 2	2	0.002	0.010
Peptide 3	3	0.011	0.015
Peptide 4	4	0.077	0.020
Peptide 5	5	0.210	0.025
Peptide 6	6	0.350	0.030
Peptide 7	7	0.410	0.035
Peptide 8	8	0.650	0.040
Peptide 9	9	0.740	0.045
Peptide 10	10	0.920	0.050

4 Notes

1. The experimental design in Fig. 2 describes a generic time course analysis of the effects of a hormone on global protein phosphorylation. The time points can be altered depending on the choice of hormone/drug as well as the system being studied. It is recommended that the length of each hormone treatment has its own time-matched control to account for fluctuations in basal phosphorylation levels with time. An alternative use of the 8plex iTRAQ methodology would be a dose-response assay to determine the effects of different concentrations of a hormone/drug on global protein phosphorylation.
2. If the sample is too diluted, you will need to concentrate the sample. We recommend a centrifugal filtration unit such as a Microcon YM-10 from Millipore.
3. We recommend using an HLB 1 cc/30 mg cartridge (WAT094225 or equivalent) which has a 1–5 mg peptide binding capacity. As gravity elution is not practical, a 5 cc syringe mounted on a luer adaptor (WAT054260) is recommended for controlled positive displacement of buffers and sample.
4. The iTRAQ 8plex Multi-plex Kit provides five 1-U tubes of each of eight different iTRAQ reagents (i.e., 113, 114, 115, 116, 117, 118, 119, and 121). Each unit can label up to 100 μg of peptide sample. Therefore, you will need all five vials of each reagent to label 500 μg of peptide sample for each experimental condition.
5. After adding the iTRAQ reagent to your sample, check that the pH is between 8.0 and 8.5 to ensure efficient labeling. Other requirements for efficient labeling include avoiding buffers with primary amines (e.g., ammonium bicarbonate and Tris), a Dissolution Buffer concentration of 120–150 mM, an organic concentration >65 %, an iTRAQ reagent concentration of 40 mM ± 5 %, and a peptide concentration of 0.5–1 mg/ml.
6. Due to the presence of negatively charged phosphate groups, phosphopeptides will not bind as strongly as unphosphorylated peptides to the negatively charged SCX resin. Thus, the majority of phosphopeptides will elute in earlier SCX fractions, while unphosphorylated peptides will tend to elute later. However, due to the presence of missed trypsin cleavages and other factors, phosphopeptides can be distributed across all SCX fractions.
7. iTRAQ-labeled peptides are larger and more hydrophobic than their unlabeled peptide counterparts. Adding 3–5 % ACN

to resuspend dried peptides following the labeling reaction may increase recovery.

8. We use the Fe-NTA Phosphopeptide Enrichment Kit, although Ga^{+3}-based IMAC or TiO_2-based enrichment are both viable alternatives. If you choose the Fe-NTA method, we recommend that the final desalting step is done using Pierce Graphite Spin Columns.
9. The LC portion of this particular LC-MS/MS setup uses a C18 pre-column for desalting. The captured peptides are then directed to a PicoFrit reversed-phase analytical column.
10. Besides Sequest, other algorithms that can be used to search phosphoproteomic data include Mascot, InsPecT, and X!Tandem. Also, besides searching the RefSeq protein database, other protein databases (e.g., Swiss-Prot) can be used.
11. As phosphopeptides often contain multiple serine, threonine, and tyrosine residues, it is of critical importance to verify that the site(s) of phosphorylation reported by the initial search algorithm are correct or if an alternative phosphorylation configuration is more likely. Search engines such as Sequest are not designed for this purpose and often report incorrect phosphorylation sites.
12. This method ensures that the more intense spectra (i.e., the ones that often have more accurate reporter ion intensities) contribute more to the final calculated ratio.
13. We recommend replicating each experimental condition at least three times (biological replicates are preferable to technical replicates) to obtain the most accurate quantitation values and for proper statistical analyses.
14. Other multiple testing correction methods include Bonferroni, Bonferroni Step-Down, and Westfall and Young Permutation. These methods are more stringent (i.e., they will produce a lower number of false positives and a higher number of false negatives) which will reduce the sensitivity of the analysis.

Acknowledgments

This work was supported by the Intramural Programs of the National Heart, Lung, and Blood Institute (Project Z01-HL-001285) and by the National Research University Project, Office of Higher Education Commission (WCU-006-HR-57).

References

1. Nuwaysir LM, Stults JT (1993) Electrospray ionization mass spectrometry of phosphopeptides isolated by on-line immobilized metal-ion affinity chromatography. J Am Soc Mass Spectrom 4(8):662–669
2. Kange R, Selditz U, Granberg M, Lindberg U, Ekstrand G, Ek B, Gustafsson M (2005) Comparison of different IMAC techniques used for enrichment of phosphorylated peptides. J Biomol Tech 16(2):91–103
3. Thingholm TE, Jensen ON (2009) Enrichment and characterization of phosphopeptides by immobilized metal affinity chromatography (IMAC) and mass spectrometry. Methods Mol Biol 527:47–56, Xi
4. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4(7):873–886
5. Klemm C, Otto S, Wolf C, Haseloff RF, Beyermann M, Krause E (2006) Evaluation of the titanium dioxide approach for MS analysis of phosphopeptides. J Mass Spectrom 41(12):1623–1632
6. Schlosser A, Pipkorn R, Bossemeyer D, Lehmann WD (2001) Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. Anal Chem 73(2):170–176
7. Bateman RH, Carruthers R, Hoyes JB, Jones C, Langridge JI, Millar A, Vissers JP (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. J Am Soc Mass Spectrom 13(7):792–803
8. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. Anal Chem 76(13):3590–3598
9. Stensballe A, Jensen ON, Olsen JV, Haselmann KF, Zubarev RA (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. Rapid Commun Mass Spectrom 14(19):1793–1800
10. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci U S A 104(7):2199–2204
11. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP (2011) Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Mol Cell Proteomics 10(12):M111
12. Altelaar AF, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HT, Heck AJ, Mohammed S (2013) Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics 88:14–26
13. Dephoure N, Gygi SP (2012) Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. Sci Signal 5(217):rs2
14. Everley RA, Kunz RC, McAllister FE, Gygi SP (2013) Increasing throughput in targeted proteomics assays: 54-plex quantitation in a single mass spectrometry run. Anal Chem 85(11):5340–5346
15. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75(8):1895–1904
16. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3(12):1154–1169
17. Pichler P, Kocher T, Holzmann J, Mazanek M, Taus T, Ammerer G, Mechtler K (2010) Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. Anal Chem 82(15):6549–6558
18. Pottiez G, Wiederin J, Fox HS, Ciborowski P (2012) Comparison of 4-plex to 8-plex iTRAQ quantitative measurements of proteins in human plasma samples. J Proteome Res 11(7):3774–3781
19. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat Methods 8(11):933–935
20. Ting L, Rad R, Gygi SP, Haas W (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 3(11):937–940
21. Viner R, Bomgarden R, Blank M, Rogers J (2013) Increasing the Multiplexing of Protein

Quantitation from 6- to 10-Plex with Reporter Ion Isotopologues. Thermo Scientific Poster Note PN ASMS13, W617

22. Hoffert JD, Pisitkun T, Saeed F, Song JH, Chou CL, Knepper MA (2012) Dynamics of the G protein-coupled vasopressin V2 receptor signaling network revealed by quantitative phosphoproteomics. Mol Cell Proteomics 11(2):M111
23. Diedrich JK, Pinto AF, Yates JR III (2013) Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. J Am Soc Mass Spectrom 24(11):1690–1699
24. Saeed F, Pisitkun T, Hoffert JD, Wang G, Gucek M, Knepper MA (2012) An efficient dynamic programming algorithm for phosphorylation site assignment of large-scale mass spectrometry data. Proceedings (IEEE Int Conf Bioinformatics Biomed), pp 618–625
25. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 24(10):1285–1292
26. Slotta DJ, McFarland MA, Markey SP (2010) MassSieve: panning MS/MS peptide data for proteins. Proteomics 10(16):3035–3039
27. Tchapyjnikov D, Li Y, Pisitkun T, Hoffert JD, Yu MJ, Knepper MA (2010) Proteomic profiling of nuclei from native renal inner medullary collecting duct cells using LC-MS/MS. Physiol Genomics 40(3):167–183
28. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 57:289–300

Chapter 5

Identification of Direct Kinase Substrates Using Analogue-Sensitive Alleles

Daniel A. Rothenberg, Elizabeth A. Gordon, Forest M. White, and Sebastian Lourido

Abstract

Identifying the substrates of protein kinases remains a major obstacle in the elucidation of eukaryotic signaling pathways. Promiscuity among kinases and their substrates coupled with the extraordinary plasticity of phosphorylation networks renders traditional genetic approaches or small-molecule inhibitors problematic when trying to determine the direct substrates of an individual kinase. Here we describe methods to label, enrich, and identify the direct substrates of analogue-sensitive kinases by exploiting their steric complementarity to artificial ATP analogues. Using calcium-dependent protein kinases of *Toxoplasma gondii* as a model for these approaches, this protocol brings together numerous advances that enable labeling of kinase targets in semi-permeabilized cells, quantification of direct labeling over background, and highly specific enrichment of targeted phosphopeptides.

Key words AS kinase, SILAC, IMAC, LC MS/MS, Toxoplasmosis, Quantitative analysis

1 Introduction

Traditionally, kinase-substrate interactions have been determined using in vitro kinase assays consisting of purified or recombinant kinases, their substrates, and cofactors necessary for activity. These kinase assays are typically limited to one putative kinase-substrate interaction at a time, require a priori knowledge, and therefore preclude the discovery of novel substrates on a large scale. Through the use of protein microarrays, it is now possible to perform kinase assays on thousands of proteins simultaneously, which has led to the identification of dozens of novel putative substrates for selected kinases [1]. Unfortunately, in vitro kinase assays, in either traditional or microarray format, lose the specificity conferred by cellular context, including subcellular localization, interaction with other kinases and phosphatases, and many other factors that constrain kinase activity and give rise to the regulatory networks observed in vivo.

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_5,

In contrast, various phosphoproteomic approaches have been developed to enrich for phosphopeptides from whole-cell lysates, in an attempt to characterize the signaling states of living cells. Among these techniques, immobilized metal affinity chromatography (IMAC) and metal-oxide affinity chromatography (MOAC; e.g., TiO_2) have functioned as robust and effective methods for enriching phosphopeptides, which are typically found at sub-stoichiometric abundance compared to their non-phosphorylated counterparts [2, 3]. These mass spectrometry (MS)-based phosphoproteomics approaches have enabled the rapid, in-depth investigation of protein phosphorylation-mediated signaling networks in a diverse array of biological systems. One key advantage of these MS-based techniques is the identification of the exact phosphorylation sites on a given protein, thereby enabling functional investigation of each site through site-directed mutagenesis. Moreover, site specificity allows for analysis of the dataset to identify motifs enriched in the sequences surrounding the phosphorylation sites. When compared against databases of known kinase motifs, this latter information can function as a fingerprint of protein kinase activity in the biological sample. Additionally, phosphoproteomics can reveal the network-wide effects of manipulating the expression or activation state of a given protein kinase, including altered phosphorylation sites on direct kinase substrates and indirect effects resulting from downstream pathways and feedback networks. The limitations of these global approaches lie in the complexity of the phosphoproteome, which frequently obscures low-abundance peptides and precludes identification of the kinase(s) responsible for a specific phosphorylation event.

To address these challenges, chemical-genetic approaches pioneered by Kevan Shokat's lab distinguish the activity of specific kinases through genetic manipulation of the kinase to engineer steric complementarity to synthetic ATP analogues [4]. The large hydrophobic residue(s) occupying the gatekeeper position in the nucleotide-binding pocket of most kinases provide strict selectivity for ATP. Mutation of the gatekeeper to a glycine or alanine preserves activity in most kinases while providing complementarity to N^6-substituted (bulky) ATP analogues. Kinases with expanded ATP-binding pockets have been termed analogue sensitive (AS) given their susceptibility to bulky pyrazolo [3,4-d] pyrimidine (PP) derivatives [5]. Most recently, bulky ATPγS analogues, containing a γ-thiophosphate as well as the N^6-substituent, have been used in conjunction with AS-kinase alleles. This combination leads to the thiophosphorylation of substrates by the kinase of interest, thereby enabling their subsequent enrichment through thiol-specific chemistry [6–8].

This protocol focuses on the calcium-dependent protein kinases (CDPKs) of the parasite *Toxoplasma gondii* as a test case for the identification of kinase substrates through AS-kinase alleles.

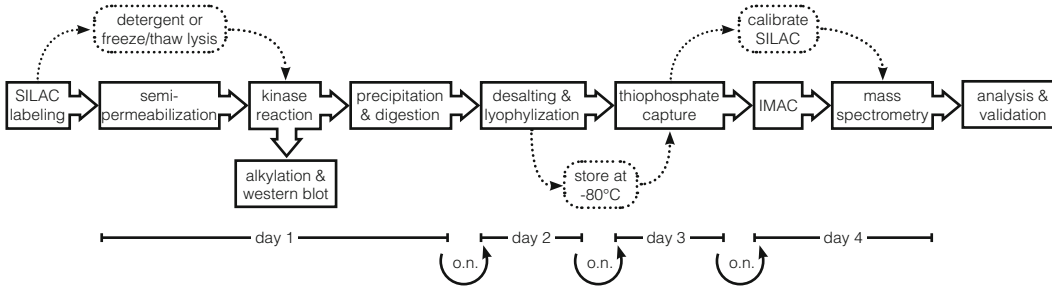


Fig. 1 Workflow for the identification of AS-kinase substrates through thiophosphorylation. Alternative steps are illustrated with *dotted lines*. A general timeline for the experiments is shown below the diagram; o.n., overnight incubation

T. gondii is a widely distributed apicomplexan parasite that infects nearly a quarter of the world's population, although its pathogenesis is mainly restricted to immune-compromised individuals and the developing fetus [9]. Due to its ease of culture and genetic tractability, *T. gondii* has emerged as a model system for the phylum that includes the etiological agents of malaria. CDPKs have been shown to be essential regulators of parasite development, motility, and invasion [10]. Because they are absent from the mammalian host, CDPKs have also garnered attention as potential therapeutic targets [11–14]. In *T. gondii* one of these kinases, TgCDPK1, has a small gatekeeper (Gly^{128}) and is therefore naturally susceptible to bulky PP derivatives [11, 12, 15] and able to utilize bulky ATPγS analogues, like 6-Fu-ATPγS [KTPγS; 16]. This gatekeeper can be mutated to a larger residue ($Gly^{128}Met$), which preserves kinase activity but renders the kinases insensitive to the inhibitors and unable to utilize the bulky artificial substrates. Comparison of the wild-type and mutant strains has been previously used to identify proteins thiophosphorylated by TgCDPK1 in parasite lysates [16]. In the current protocol, we combine this approach with a novel semi-permeabilization strategy that utilizes the pore-forming toxin aerolysin to introduce KTPγS into parasites, without disrupting the cellular ultrastructure. Additionally, through the use of stable isotope labeling with amino acids in cell culture [SILAC; 17], we can quantitatively compare thiophosphorylation in strains bearing the different kinase alleles and thereby distinguish TgCDPK1 targets from background utilization of the analogue. We also describe highly effective methods of enriching for thiophosphorylated peptides, including in-column elution and cleanup, and immobilized metal ion affinity chromatography (IMAC), which coupled with mass spectrometry provides a sensitive means of identifying the thiophosphorylated proteins and the exact modified residue(s) [7]. The workflow for the indicated protocol is illustrated in Fig. 1.

These methods are generalizable to a variety of cell-based systems, and we have attempted to highlight key steps that should be

optimized for other kinases and other organisms. The approaches described below will enable the discovery of new substrates and account for the effects of subcellular localization on kinase function and specificity.

2 Materials

2.1 Cell Culture

1. SILAC media: DMEM for SILAC supplemented with 0.1 mg/ml l-lysine [$^{12}C_6{}^{14}N_2$ (light), $^{13}C_6{}^{14}N_2$ (medium) or $^{13}C_6{}^{15}N_2$ (heavy)], 0.1 mg/ml l-arginine [$^{12}C_6{}^{14}N_4$ (light), $^{13}C_6{}^{14}N_4$ (medium) or $^{13}C_6{}^{15}N_4$ (heavy)], 0.23 mg/ml l-proline, 10 % dialyzed FBS for SILAC.
2. Human foreskin fibroblasts (SCRC-1041; ATCC, Manassas, VA) grown in DMEM supplemented with 10 % FBS.
3. T12.5 and T175 tissue culture flasks.
4. Dulbecco's phosphate-buffered saline (PBS).
5. Intracellular buffer (IB; 10×): 1.37 M KCl, 50 mM NaCl, 100 mM $MgCl_2$, 200 mM HEPES (pH 7.4). Diluted tenfold in MilliQ water for use with cells (1× IB).

2.2 Semi-permeabilization and Thiophosphorylation

1. Aerolysin: HIS-tagged proaerolysin, purified as previously described [18], and stored in 50 mM Tris (pH 7.5), 300 mM NaCl, 10 % (v/v) glycerol at –80 °C. Dilute to 6 μg/ml in IB prior to use.
2. Kinase buffer: 1× IB, 4 mM CaEGTA (*see* **Note 1**), 1 mM GTP, 100 μM ATP, 10 μM KTPγS (6-Fu-ATPγS; BioLog, Bremen, Germany), 1× Halt protease/phosphatase inhibitor cocktail.

2.3 Lysis, Protein Precipitation, and Digestion

1. Lysis buffer: 1× IB, 1 % Triton X-100, 10 mM EGTA, 1× Halt protease/phosphatase inhibitor cocktail.
2. DC Protein Assay.
3. Methanol.
4. Chloroform.
5. 8 M urea dissolved in water.
6. 1 M TCEP solution in water. Store single-use aliquots at –80 °C, for up to 6 months.
7. Digest buffer: 100 mM ammonium acetate (pH 8.9), 1 mM $CaCl_2$. Add 2 mM TCEP immediately before digest.
8. Trypsin Gold, mass spectrometry grade.

2.4 Analysis of Thiophosphorylation

1. 50 mM p-nitrobenzyl mesylate.
2. Reagents and equipment for SDS-PAGE and Western blotting.

3. Anti-thiophosphate ester antibody (ab92570, Abcam, Cambridge, MA).

2.5 Desalting and Lyophilization

1. C18 Sep-Pak Cartridges.
2. Glacial acetic acid.
3. 0.1 % acetic acid.
4. 90 % acetonitrile/0.1 % acetic acid.
5. 50 % acetonitrile/0.1 % acetic acid.
6. Vacuum centrifuge.

2.6 Preparation of Large Fritted Column

1. 530 μm i.d. 700 μm o.d. fused silica capillary.
2. Kasil 1 (potassium silicate solution).
3. Formamide.
4. Teflon tubing (PTFE 0.011 inch i.d., 0.0625 o.d.)
5. Hand drill with 72-gauge bit.

2.7 Preparation of R2 Column

1. 100 μm i.d. 360 μm o.d. fused silica capillary.
2. POROS R2 10 μm beads.
3. Small magnetic stir bar.
4. Angiotensin peptide diluted in 0.1 % acetic acid.

2.8 Thiophosphate Capture

1. Thiophosphate binding buffer: 25 mM HEPES (pH 7.0), 50 % acetonitrile, 2 mM TCEP (added immediately before use).
2. SulfoLink Coupling Resin.
3. Bovine serum albumin (BSA).

2.9 Thiophosphate Cleaning and Elution

1. SulfoLink Quenching Buffer: 25 mM HEPES (pH 8.5), 50 % acetonitrile, 5 mM DTT (added immediately before use).
2. 5 % formic acid.
3. Oxidizing elution buffer: 2 mg/ml potassium monopersulfate, prepared immediately before use.
4. Tabletop centrifuge.
5. Sample rotator.

2.10 Immobilized Metal Affinity Chromatography Enrichment

1. Immobilized metal affinity chromatography (IMAC) column: *see* ref. 20 for details on preparing and testing IMAC columns.
2. EDTA rinse: 100 mM EDTA (pH 8.9).
3. Iron chloride: 100 mM iron (III) chloride.
4. 70 % acetonitrile/0.2 M acetic acid.
5. Organic rinse: 25 % acetonitrile, 1 % acetic acid, 100 mM NaCl.

6. Autosampler vials.
7. IMAC elution buffer: 250 mM NaH_2PO_4 (pH 8.9).

2.11 Liquid Chromatography and Mass Spectrometry

1. HPLC aqueous solvent: 0.2 M acetic acid in ultrapure water.
2. HPLC organic solvent: 70 % acetonitrile, 0.2 M acetic acid in ultrapure water.
3. Thermo Scientific Easy-nLC 100 in conjunction with a Thermo Scientific Q-Exactive ion trap mass spectrometer.
4. MASCOT Distiller version 2.5 in conjunction with MASCOT Server version 2.4 (Matrix Science, Boston, MA).
5. Computer-assisted manual validation (CAMV) [21].

3 Methods

3.1 Cell Culture

1. *T. gondii* $CDPK1^G$ and $CDPK1^M$ [22] are grown on a 2-day cycle in confluent monolayers of human foreskin fibroblasts. To label, grow parasites for three passages in the presence of the appropriate SILAC media. This is achieved by rinsing the recipient monolayer with PBS, before adding the SILAC media and sufficient parasites to achieve complete lysis of the monolayer in 2 days. For the final passage, inoculate two T175 flasks per strain.
2. Harvest the parasites by resuspending the media from fully lysed T175 flasks, and removing cell debris by filtration through 3 μm filters. Rinse filters with cold 1× IB.
3. Pellet parasites by centrifugation, 10 min, 400 × *g*, 4 °C, and resuspend each strain in 10 ml 1× ICB. Pellet again and resuspend each strain in 400 μl 1× ICB (*see* **Note 2**).

3.2 Semi-permeabilization and Thiophosphorylation

1. Add 400 μl of aerolysin solution to each sample and incubate for 10 min at 37 °C (*see* **Note 3**).
2. Add 800 μl of kinase buffer to each sample and incubate for 30 min at 30 °C.
3. Pellet parasites by centrifugation, 10 min, 400 × *g*, 4 °C, and remove supernatant. Keep pellets on ice and immediately proceed to lysis and protein precipitation.

3.3 Lysis, Protein Precipitation, and Digestion

1. Add 250 μl of cold lysis buffer to each sample and resuspend the parasite pellets. Remove 5 μl of each sample for analysis (continued in Subheading 3.4).
2. Use 5 μl of each sample for protein quantitation using the DC protein assay following the manufacturer's instructions.
3. To each sample, add 800 μl methanol and 200 μl chloroform and vortex.

4. Add 600 µl dH_2O, vortex, and centrifuge at maximum speed, 5 min, 4 °C.
5. Remove the upper (aqueous) layer, without disrupting the protein precipitate in the interphase.
6. Add 375 µl methanol, vortex, and centrifuge at maximum speed, 15 min, 4 °C.
7. Carefully remove liquid and allow protein to air-dry for 15 min.
8. Resuspend each pellet in 200 µl 8 M urea, and combine equal amounts of each sample according to the protein quantitation (*see* **Note 4**).
9. Add 5 volumes (~2 ml) of trypsin digest buffer and 20 µg of Trypsin Gold per mg of protein. Digest rotating overnight, at room temperature.

3.4 Analysis of Thiophosphorylation

1. For analysis, add 35 µl of 1× IB to each sample and 1.6 µl 50 mM PNBM (2 mM final).
2. Incubate for 2 h at room temperature, before resolving samples by SDS-PAGE.
3. Blot and probe with rabbit-anti-thiophosphate ester antibody and an appropriate loading control.

3.5 Desalting and Lyophilization

1. Acidify the digested sample with glacial acetic acid to 10 %. Spin down and remove any debris that may be present to prevent clogging of the Sep-Pak cartridges.
2. Obtain C18 Sep-Pak cartridges. Sep-Pak light cartridges may be used for samples up to 1 mg total protein. Sep-Pak plus cartridges should be used for samples up to 4 mg total protein.
3. Wash the Sep-Pak cartridge with 10 ml 0.1 % acetic acid at a flow rate of 2 ml/min.
4. Equilibrate the Sep-Pak cartridges with 10 ml 90 % acetonitrile/0.1 % acetic acid at a flow rate of 2 ml/min.
5. Wash the Sep-Pak cartridges with 10 ml 0.1 % acetic acid at a flow rate of 2 ml/min.
6. Load the digested samples at a flow rate of 0.5 ml/min.
7. Wash the loaded Sep-Pak cartridges with 10 ml 0.1 % acetic acid at a flow rate of 2 ml/min.
8. Elute the peptides into a clean conical tube with 5 ml 50 % acetonitrile/0.1 % acetic acid at a flow rate of 1 ml/min.
9. Reduce the total volume of each sample to less than 1 ml in a vacuum centrifuge.
10. Snap freeze the sample by immersing in liquid nitrogen for 10 min.

11. Lyophilize the sample overnight, refreezing the sample if it thaws during lyophilization (*see* **Note 5**).

3.6 Thiophosphate Capture

1. Resuspend the lyophilized peptides in an appropriate volume of thiophosphate-binding buffer supplemented with 25 μg/ml bovine serum albumin (BSA). The final pH of the peptides should be ~5.5 (*see* **Note 6**).
2. Transfer SulfoLink bead slurry to a microcentrifuge tube. 25 μL of bead slurry should be used per 1 mg of protein in the original lysate, up to 100 μl of slurry (*see* **Note 7**).
3. Wash the SulfoLink beads twice in 500 μl binding buffer, and once with binding buffer supplemented with 25 μg/ml BSA, each time for 5 min rotating in the dark. Pellet beads by centrifugation at 100×*g* for 10 s and carefully remove supernatant.
4. Apply resuspended peptides to SulfoLink beads and allow thiophosphate capture to take place overnight rotating at room temperature in the dark.

3.7 Preparation of Large Fritted Capillary

1. Cut 35 cm length of 530 μm i.d. fused silica capillary.
2. Combine 80 μl Kasil and 20 μl formamide, vortex, and centrifuge for 30 s at maximum speed.
3. Draw up 1–2 cm of material from the center of the frit solution by capillary action.
4. Bake frit in oven at 80 °C for 15 min to ensure complete polymerization.
5. Wash frit by flowing acetonitrile over frit at very low pressure to prevent spraying out of column.
6. Using a hand drill with 72-gauge bit, drill out one side of a 2.5 cm piece of Teflon tubing to form a seal with the 530 μm capillary. Remove the debris and widen the other side of the connector by pushing 360 μm OD fused silica capillary through the connector.
7. Attach the connector to the fritted end of the capillary (*see* **Note 8**).

3.8 Preparation of R2 Column

1. Cut a 25 cm length of 100 μm i.d. capillary
2. Combine 100 μl Kasil and 20 μL formamide, vortex, and centrifuge for 30 s at maximum speed.
3. Draw up 1 cm of solution by capillary action.
4. Bake frit in oven at 80 °C for 15 min to ensure complete polymerization.
5. Wash by flowing acetonitrile over frit at very low pressure.

6. Suspend POROS R2 beads in 80 % acetonitrile/20 % isopropanol with a magnetic stir bar.
7. Load POROS R2 beads into capillary at 500 psi until 10 cm of beads are packed.
8. Wash column with 0.1 % acetic acid for 10 min at 100–200 psi.
9. Load 5 pmol angiotensin peptide in 0.1 % acetic acid.
10. Place column on HPLC and run a 20-min gradient from 0 to 70 % acetonitrile/0.2 M acetic acid to condition the column.
11. Run another gradient over the column immediately before use.

3.9 Thiophosphate Cleaning and Elution

1. Pellet the beads by centrifugation for 30 s at 100×*g*, and remove supernatant (*see* **Note 9**).
2. Wash beads with 500 µl binding buffer twice for 5 min rotating in the dark.
3. Wash beads with 500 µl quenching buffer for 5 min rotating in the dark (*see* **Note 10**).
4. Wash beads with 500 µl binding buffer for 5 min rotating.
5. Wash beads with 500 µl 5 % formic acid, incubating for 5 min without rotating (*see* **Note 11**).
6. Wash the beads with 500 µl binding buffer for 5 min rotating.
7. Resuspend the beads in 1 ml binding buffer and transfer to a 2 ml glass vial with a magnetic stir bar. Use a magnetic stir plate to keep the beads in suspension for subsequent steps.
8. Use helium pressure device to load the beads into the large fritted capillary (*see* Subheading 3.7), periodically adding binding buffer to prevent the beads from running dry. Continue until more than 90 % of the beads have been loaded. Reduce the dead space in the capillary by trimming to 3 cm below the end of the bead pack.
9. Wash the beads with 0.1 % acetic acid for 5 min at a flow rate of 200 µl/min.
10. Attach R2 column (*see* Subheading 3.8) to the SulfoLink column using the Teflon connector.
11. Adjust the pressure to achieve a flow rate of 4 µl/min using 0.1 % acetic acid.
12. Replace the 0.1 % acetic acid with the oxidizing elution buffer for 10 min to elute the peptides onto the R2 column.
13. Replace the oxidizing elution buffer with 0.1 % acetic acid and wash the column for at least 25 min (*see* **Note 12**).
14. Remove the R2 column from the Teflon connector and discard the SulfoLink column. Rinse the R2 column with 0.1 % acetic acid for 5 min at 150 psi to remove residual Oxone prior to IMAC enrichment.

3.10 IMAC Enrichment

1. Prepare an IMAC column for metal affinity enrichment of phosphopeptides. The preparation and testing of IMAC columns are discussed in detail in ref. 20.
2. Rinse the IMAC column with EDTA for 10 min at a flow rate of 12 µl/min.
3. Wash the IMAC column with MilliQ water for 10 min at a flow rate of 12 µl/min.
4. Load the IMAC column with iron chloride at a rate of 12 µl/min for 30 min (*see* **Note 13**).
5. Rinse the IMAC column with 0.1 % acetic acid for 10 min at a flow rate of 12 µl/min.
6. Attach the bottom of the IMAC column to the top of the R2 column using a Teflon connector. Using 0.1 % acetic acid, adjust the flow rate to less than 1 µl/min.
7. Replace 0.1 % acetic acid with 70 % acetonitrile/0.2 M acetic acid and elute the peptides onto the IMAC column with a total volume of 10 µl (*see* **Note 14**). Discard the R2 column.
8. Wash the IMAC column with organic rinse for 10 min at a flow rate of 12 µl/min.
9. Wash the column for 10 min with 0.1 % acetic acid at a flow rate of 12 µl/min.
10. Obtain a fresh autosampler vial and place it inverted on top of the IMAC column (*see* **Note 15**).
11. Elute phosphopeptides into the autosampler vial with 40 µl of IMAC elution buffer at a flow rate of 4 µl/min (*see* **Note 16**).

3.11 Liquid Chromatography and Mass Spectrometry

1. Analyze peptides eluted from the IMAC column by LC-MS/MS using reverse-phase chromatography performed in-line with a Q Exactive mass spectrometer (*see* **Note 17**). To prevent phosphate salt ions from contaminating the sample, the pre-column should be washed extensively with 0.2 M acetic acid following sample loading.
2. Elute peptides using a 120-min gradient (0–70 % acetonitrile in 0.2 M acetic acid).
3. Acquire data using the mass spectrometer in data-dependent acquisition mode. Typical settings include one full-scan mass spectrum followed by ten tandem mass spectra, using an intensity threshold of 3.3×10^3. Automatic gain control target value and maximum fill times are typically set to 3×10^6 and 50 ms for the full-scan mass spectra, and 1×10^5 and 300 ms for the tandem mass spectra.

3.12 Data Analysis

1. Load MS data file into MASCOT Distiller and search using MASCOT Server using the following search parameters: type of search, MS/MS Ion Search; quantitation, SILAC K+6 R+6

[MD] (2 component SILAC); enzyme, trypsin; variable modifications, dioxidation (M), oxidation (M), phospho (STY), label, 13C(6) (K), label, 13C(6) (R); mass values, monoisotopic; protein mass, unrestricted; peptide mass tolerance, ± 10 ppm; fragment mass tolerance, ± 0.8 Da; max missed cleavages, 2; instrument type, ESI-FTICR.

2. All tandem mass spectra of potential kinase substrates should be manually validated to confirm identity of the phosphorylated peptide and phosphorylation site, using a minimum MASCOT score cutoff of 25 [26]. Use CAMV to facilitate this process [21]. Use the SILAC ratios to establish a threshold for determining kinase substrates (*see* **Note 18**).

4 Notes

1. CaEGTA is prepared as previously described [19] and under the given reaction conditions should achieve ~20 μM free Ca^{2+}.
2. Note that some kinases will require stimulation of the cells prior to harvesting to achieve full activity, as is the case with mammalian ERK2 [7]. Alternatively, cells can by lysed via two freeze/thaw cycles. Using this protocol, cells are scraped in residual PBS following aspiration and collected into a microcentrifuge tube. The cells are alternated between 1 min submerged in liquid nitrogen and thawing in a 56 °C water bath. The lysate can then be diluted in kinase buffer. Freeze/thaw lysis removes the need for methanol/chloroform extraction later in the protocol.
3. It is also possible to add aerolysin to cells at 4 °C, wash away the unbound toxin, and then allow permeabilization to occur at 37 °C.
4. Typical yields from two completely lysed T175 flasks range between 1 and 1.5 mg of total protein.
5. Lyophilized peptides may be stored at −80 °C for several months.
6. The pK_a of *O*-methyl thiophosphate is 1.67, while the pK_a of cysteines ranges from 8.0 to 8.3 depending on the local environment. Multiple groups have reported that performing the thiophosphate capture at an acidic pH increases selectivity for thiophosphate in the presence of free cysteines [23, 24]. However, we have not been able to observe a pH-dependent change in selectivity between thiophosphate and cysteine capture.
7. The SulfoLink beads consist of light and air-sensitive iodoacetyl groups conjugated to an agarose scaffold. Light exposure should be limited by wrapping tubes in aluminum foil until the quenching step.

8. Optional: To test the integrity of the frit, following the acetonitrile wash, suspend fresh SulfoLink beads in binding buffer and load into the 530 μm capillary and look for packing near the frit. Once bead packing is observed, flip the column upside down and use acetonitrile to push the beads out of the column.
9. Analyze a dilution of the supernatant by mass spectrometry to establish the ratio at which the different strains were combined by comparing the SILAC ratios of diagnostic abundant peptides, as previously performed for iTRAQ-labeled samples [25].
10. From this point onward, the beads no longer have to be treated as light sensitive since the free iodoacetyl groups will have been quenched.
11. Do not rotate during this incubation since it will decrease the pelleting efficiency of the beads.
12. It is important to flush the entire length of the column to wash all the eluted peptides into the R2 column. The length of this wash can be adjusted as necessary based on the length of the SulfoLink column.
13. Optionally, reverse the direction of the IMAC column and load with iron chloride for an additional 10 min.
14. A slow flow rate is critical to ensure that phosphopeptides eluted from the R2 column have time to bind to the IMAC column.
15. The elution buffer will collect at the bottom of the vial. If you are worried about spilling your sample, you may attach a short piece of empty capillary to the top of the IMAC column and use that to direct the elution to an upright autosampler vial.
16. The IMAC column can be rinsed with MilliQ water or stored as is.
17. The LC-MS/MS protocol will depend on your particular equipment and configuration. We recommend consulting an experienced mass spectrometry facility about how to analyze your sample.
18. We set our threshold at three standard deviations above the mean for non-phosphorylated peptides.

References

1. Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, Smith D, Gerstein M, Reed MA, Snyder M (2000) Analysis of yeast protein kinases using protein chips. Nat Genet 26:283–289. doi:10.1038/81576
2. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJD (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4:873–886. doi:10.1074/mcp.T500007-MCP200
3. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis

by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol 20:301–305. doi:10.1038/nbt0302-301

4. Shah K, Liu Y, Deirmengian C, Shokat KM (1997) Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. Proc Natl Acad Sci U S A 94:3565–3570
5. Bishop A, Kung C, Shah K, Witucki L, Shokat K, Liu Y (1999) Generation of monospecific nanomolar tyrosine kinase inhibitors via a chemical genetic approach. J Am Chem Soc 121:627–631
6. Allen JJ, Li M, Brinkworth CS, Paulson JL, Wang D, Hübner A, Chou W-H, Davis RJ, Burlingame AL, Messing RO, Katayama CD, Hedrick SM, Shokat KM (2007) A semisynthetic epitope for kinase substrates. Nat Methods 4:511–516. doi:10.1038/nmeth1048
7. Carlson SM, Chouinard CR, Labadorf A, Lam CJ, Schmelzle K, Fraenkel E, White FM (2011) Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. Sci Signal 4:rs11. doi:10.1126/scisignal.2002010
8. Hertz NT, Wang BT, Allen JJ, Zhang C, Dar AC, Burlingame AL, Shokat KM (2010) Chemical genetic approach for kinase-substrate mapping by covalent capture of thiophosphopeptides and analysis by mass spectrometry. Curr Protoc Chem Biol 2:15–36
9. Montoya JG, Liesenfeld O (2004) Toxoplasmosis. Lancet 363:1965–1976. doi:10.1016/S0140-6736(04)16412-X
10. Billker O, Lourido S, Sibley LD (2009) Calcium-dependent signaling and kinases in apicomplexan parasites. Cell Host Microbe 5:612–622. doi:10.1016/j.chom.2009.05.017
11. Lourido S, Shuman J, Zhang C, Shokat KM, Hui R, Sibley LD (2010) Calcium-dependent protein kinase 1 is an essential regulator of exocytosis in Toxoplasma. Nature 465:359–362. doi:10.1038/nature09022
12. Ojo KK, Larson ET, Keyloun KR, Castaneda LJ, Derocher AE, Inampudi KK, Kim JE, Arakaki TL, Murphy RC, Zhang L, Napuli AJ, Maly DJ, Verlinde CLMJ, Buckner FS, Parsons M, Hol WGJ, Merritt EA, van Voorhis WC (2010) Toxoplasma gondii calcium-dependent protein kinase 1 is a target for selective kinase inhibitors. Nat Struct Mol Biol 17:602–607. doi:10.1038/nsmb.1818
13. Murphy RC, Ojo KK, Larson ET, Castellanos-Gonzalez A, Perera BGK, Keyloun KR, Kim JE, Bhandari JG, Muller NR, Verlinde CLMJ, White AC, Merritt EA, van Voorhis WC, Maly DJ (2010) Discovery of potent and selective inhibitors of calcium-dependent protein kinase 1 (CDPK1) from C. parvum and T. gondii. ACS Med Chem Lett 1:331–335. doi:10.1021/ml100096t
14. Kato N, Sakata T, Breton G, Le Roch KG, Nagle A, Andersen C, Bursulaya B, Henson K, Johnson J, Kumar KA, Marr F, Mason D, McNamara C, Plouffe D, Ramachandran V, Spooner M, Tuntland T, Zhou Y, Peters EC, Chatterjee A, Schultz PG, Ward GE, Gray N, Harper J, Winzeler EA (2008) Gene expression signatures and small-molecule compounds link a protein kinase to Plasmodium falciparum motility. Nat Chem Biol 4:347–356. doi:10.1038/nchembio.87
15. Sugi T, Kato K, Kobayashi K, Kurokawa H, Takemae H, Gong H, Recuenco FC, Iwanaga T, Horimoto T, Akashi H (2011) 1NM-PP1 treatment of mice infected with Toxoplasma gondii. J Vet Med Sci 73:1377–1379
16. Lourido S, Jeschke GR, Turk BE, Sibley LD (2013) Exploiting the unique ATP-binding pocket of toxoplasma calcium-dependent protein kinase 1 to identify its substrates. ACS Chem Biol. doi:10.1021/cb400115y
17. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376–386
18. Iacovache I, Paumard P, Scheib H, Lesieur C, Sakai N, Matile S, Parker MW, van der Goot FG (2006) A rivet model for channel formation by aerolysin-like pore-forming toxins. EMBO J 25:457–466. doi:10.1038/sj.emboj.7600959
19. Tsien R, Pozzan T (1989) Measurement of cytosolic free Ca2+ with quin2. Methods Enzymol 172:230–262
20. Moser K, White FM (2006) Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. J Proteome Res 5:98–104. doi:10.1021/pr0503073
21. Curran TG, Bryson BD, Reigelhaupt M, Johnson H, White FM (2013) Computer aided manual validation of mass spectrometry-based proteomic data. Methods 61:219–226. doi:10.1016/j.ymeth.2013.03.004
22. Lourido S, Tang K, Sibley LD (2012) Distinct signalling pathways control Toxoplasma egress and host-cell invasion. EMBO J. doi:10.1038/emboj.2012.299
23. Kwon SW, Kim SC, Jaunbergs J, Falck JR, Zhao Y (2003) Selective enrichment of thiophosphorylated polypeptides as a tool for the analysis of protein phosphorylation. Mol Cell Proteomics 2:242–247. doi:10.1074/mcp.M300039-MCP200

24. Lee SE, Elphick LM, Kramer HB, Jones AME, Child ES, Anderson AA, Bonnac L, Suwaki N, Kessler BM, Gouverneur V, Mann DJ (2011) The chemoselective one-step alkylation and isolation of thiophosphorylated cdk2 substrates in the presence of native cysteine. Chembiochem 12:633–640. doi:10.1002/cbic.201000528
25. Zhang C, Kenski DM, Paulson JL, Bonshtien A, Sessa G, Cross JV, Templeton DJ, Shokat KM (2005) A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. Nat Methods 2:435–441. doi:10.1038/nmeth764
26. Nichols AM, White FM (2009) Manual validation of peptide sequence and sites of tyrosine phosphorylation from MS/MS spectra. In: Lipton MS & Paša-Tolic L (eds) Methods in molecular biology. Humana Press, Totowa, NJ, pp 143–160

Chapter 6

Quantitative Analysis of Tissue Samples by Combining iTRAQ Isobaric Labeling with Selected/Multiple Reaction Monitoring (SRM/MRM)

Ryohei Narumi and Takeshi Tomonaga

Abstract

Mass spectrometry-based phosphoproteomics is an indispensible technique used in the discovery and quantification of phosphorylation events on proteins in biological samples. The application of this technique to tissue samples is especially useful for the discovery of biomarkers as well as biological studies. We herein describe the application of a large-scale phosphoproteome analysis and SRM/MRM-based quantitation to develop a strategy for the systematic discovery and validation of biomarkers using tissue samples.

Key words Phosphoproteome, iTRAQ, SRM, MRM, IMAC

1 Introduction

Advances have recently occurred in mass spectrometry-based phosphoproteomics due to improvements in both phosphopeptide enrichment [1] and isotope labeling [2, 3] technologies. Therefore, it is now possible not only to identify several thousand phosphopeptides within one large-scale analysis [4–8], but also to accurately quantify these phosphopeptides [9–12].

A common technique for phosphopeptide enrichment is immobilized metal ion affinity chromatography (IMAC), in which metal ions are chelated to nitrilotriacetic acid- or iminodiacetic acid-coated beads, thereby forming a stationary phase to which negatively charged phosphopeptides in a mobile phase can bind [1]. Phosphopeptides in a peptide mixture prepared from biological samples by enzymatic digestion show increased affinity for the IMAC resin.

Isotope labeling techniques have been classified into two groups: metabolic labeling such as SILAC (stable isotope labeling by amino acids in a cell culture) [2] and chemical labeling such as iTRAQ (isobaric tags for relative and absolute quantification) [3].

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_6,

Chemical labeling techniques are particularly useful for quantitatively comparing proteomes between tissue samples (for example obtained from human patients). Thereagainst, metabolic labeling cannot be applied to the tissue samples, in which protein synthesis does not occur preventing the replacement of the amino acids in the proteins with isotope-labeled ones [13, 14]. Moreover, large-scale phosphoproteome analyses can be performed by combining chemical labeling with the phosphopeptide enrichment techniques, and have recently been applied to the discovery of biomarkers using tissue samples [15].

On the other hand, extensive validation for tens or hundreds of biomarker candidates identified by a large-scale phosphoproteome analysis is needed prior to their application as biomarkers. A targeted proteomic approach using selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) [16] is more appropriate for the validation of these candidates than an antibody-based approach. Antibodies with sufficient specificity and sensitivity for this validation are commonly not available, especially for phosphoproteins, and the high cost and long development time required to generate high-quality reagents are limiting factors. SRM can quantify target proteins without antibodies by monitoring the ions matching the precursor-product ion pair of m/z (SRM transition) of the target peptides. SRM using stable isotope peptides was recently used to validate candidate protein biomarkers in human tissue samples [13–15].

We herein describe the application of a large-scale phosphoproteome analysis and SRM-based quantitation to develop a strategy for the systematic discovery and validation of biomarkers using tissue samples. We first identify differentially expressed phosphopeptides using IMAC coupled with the iTRAQ technique. The phosphopeptides identified are then validated by the SRM analysis. This systematic approach has enormous potential for the discovery of *bona fide* disease biomarkers (*see* Fig. 1).

2 Materials

2.1 Homogenizing Tissue Samples and Enzymatic Digestion (See Note 1)

1. Phase-transfer surfactant A (PTS-A) buffer: 50 mM ammonium bicarbonate. Dissolve 1.0 g ammonium bicarbonate in 250 mL water.
2. Phase-transfer surfactant B (PTS-B) buffer: PTS-A buffer, 12 mM sodium deoxycholate, 12 mM sodium *N*-lauroyl sarcosinate.
3. Lysis buffer: PTS-B buffer, 1× PhosSTOP phosphatase inhibitor cocktail. Prepare the lysis buffer just before the experiment, and use it for the day.

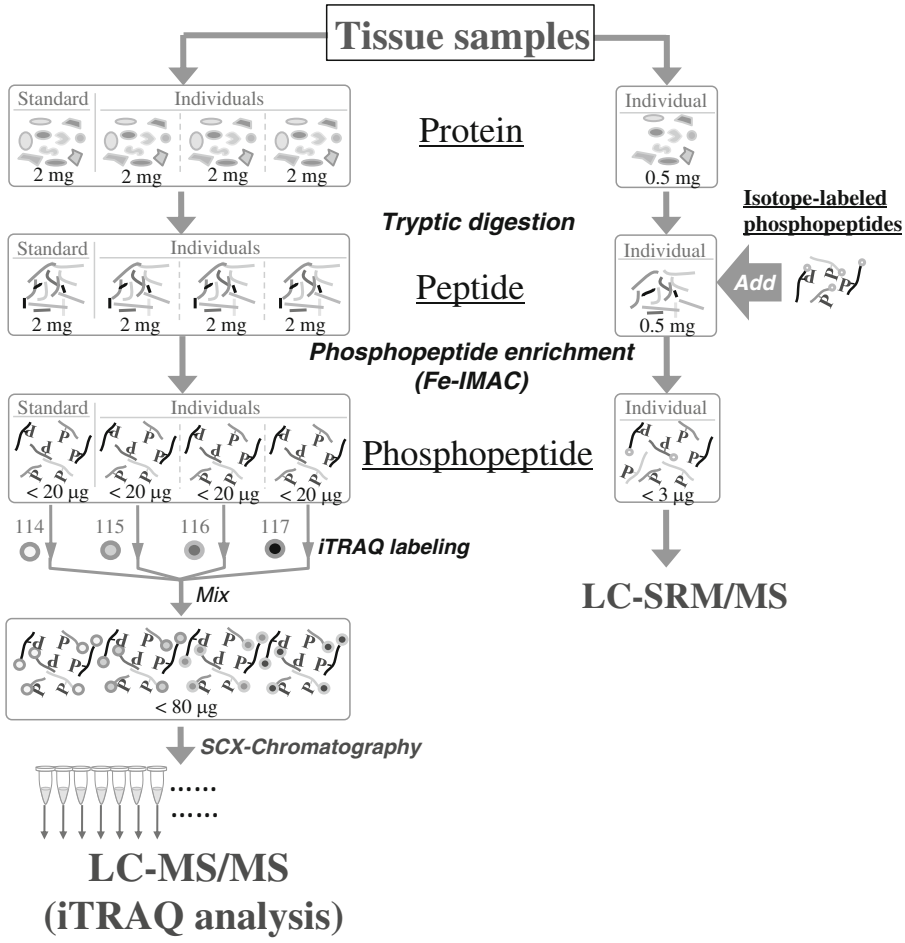


Fig. 1 Two workflows of iTRAQ analysis (*left*) and SRM analysis (*right*) are shown. In the iTRAQ analysis, Standard, which is a mixture of all analytical samples, and three individual samples are each processed into peptides and applied to Fe-IMAC to enrich the phosphopeptides. The resulting samples are labeled by iTRAQ reagents and followed by mixing the four samples. The mixture is fractionated by SCX chromatography and each fraction is analyzed by LC-MS/MS. In the SRM analysis, the individual sample is processed into peptides and followed by adding of a mixture of the isotope labeled peptides of targeted phosphorylation sites. The resulting sample is applied to Fe-IMAC to enrich the phosphopeptides and analyzed by SRM analysis

4. LysC stock solution: 1 µg/µL Lysyl endopeptidase (LysC). Dissolve lyophilized LysC in water (*see* **Note 2**). Store at −80°C.
5. Trypsin stock solution: 1 µg/µL Trypsin. Dissolve lyophilized trypsin in 10 mM HCl (*see* **Note 3**). Store at −80°C.
6. Dithiothreitol (DTT) solution (×10): 100 mM DTT. Weigh DTT powder and transfer it to a centrifuge tube. Store at 4 °C until it is used. Add 65 µL of PTS-A buffer to 1 mg of DTT and dissolve immediately prior to use.

7. Iodoacetamide (IAA) solution (×10): 500 mM IAA. Weigh IAA powder and transfer it to a centrifuge tube. Store at 4 °C in the dark until it is used. Add 10.8 μL of PTS-A buffer to 1 mg of IAA and dissolve immediately prior to use.
8. DC Protein Assay Kit.
9. Tissue grinder.
10. Liquid nitrogen.
11. Phosphate-buffered saline (PBS) buffer.
12. Bovine serum albumin (BSA), e.g., Pierce BSA Protein Assay Standards.
13. Benchtop centrifuge.
14. Sonicator, e.g., Bioruptor-UCD-250 (Cosmo Bio Japan).
15. Speed Vac.
16. 100 % Ethylacetate (sequence grade).
17. 100 % Trifluoroacetic acid (TFA) (HPLC grade).
18. 1.5 mL microtubes.

2.2 Desalting Peptide Mixtures by C18 Stage Tip (See Note 4)

1. 47 mm Empore™ C18 disk (3 M).
2. 200 μL pipet tips.
3. Methanol (LC-MS grade).
4. 80 % acetonitrile, 0.1 % TFA: Mix acetonitrile (LC-MS grade), distilled water (LC-MS grade), and TFA (HPLC grade).
5. 2 % acetonitrile, 0.1 % TFA.
6. 60 % acetonitrile, 0.1 % TFA.

2.3 IMAC

1. ProBond™ Nickel-Chelating Resin (Life Technology).
2. 50 mM EDTA-2Na in water.
3. 0.1 % acetate: Dilute acetate 1000 times in water.
4. 100 mM $FeCl_3$ in 0.1 % acetate.
5. 2 % Acetonitrile, 0.1 % TFA.
6. 60 % Acetonitrile, 0.1 % TFA.
7. 1 % Phosphate: Dilute phosphoric acid (HPLC grade, 85 %) 85 times with water.

2.4 iTRAQ Labeling

1. iTRAQ reagents: Isobaric tags for relative and absolute quantification (iTRAQ) reagents (4 plex) (Applied Biosystems).
2. 1.0 M Triethylammonium bicarbonate.
3. Ethanol.
4. pH strips.
5. Benchtop centrifuge.
6. 1.5 mL microtubes.
7. Speed Vac.

2.5 Strong Cation-Exchange Chromatography

1. Strong cation-exchange (SCX) buffer A: 25 % acetonitrile, 10 mM H_3PO_4 (pH 3). Mix 250 mL of acetonitrile (HPLC grade), approximately 650 mL of water, and 685 μL of phosphoric acid (HPLC grade, 85 %). Adjust to pH 3.0 by adding KOH solution (approximately 1.0 mL of 50 % KOH) and to 1 L with water.
2. SCX buffer B: 25 % acetonitrile, 10 mM H_3PO_4 (pH 3), 1 M KCl. Mix 250 mL of acetonitrile (HPLC grade), approximately 650 mL of water, 685 μL of phosphoric acid (HPLC grade, 85 %), and 74.55 g of KCl. Adjust to pH 3.0 by adding KOH solution (approximately 1.0 mL of 50 % KOH) and to 1 L with water.
3. HPLC system, e.g., Prominence UFLC (Shimadzu, Japan).
4. SCX column, e.g., 50 mm × 2.1 mm, 5 μm, 300 Å, ZORBAX 300SCX (Agilent Technology).
5. Sample vial/sample plate for HPLC system.
6. 1.5 mL microtubes.

2.6 iTRAQ-Based or SRM Mass Spectrometry Analysis

1. Buffer-A: 0.1 % formic acid, 2 % acetonitrile.
2. Buffer-B: 0.1 % formic acid, 90 % acetonitrile.
3. Stable isotope-labeled peptides (SI peptides) (crude grade) (Thermo Fisher Scientific) (*see* **Note 5**). Dissolve 1 μg/μL of the peptide in water. Store at −80 °C.
4. Sample vial/sample plate for MS analysis.
5. Mass spectrometer for iTRAQ analysis, e.g., LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific).
6. Mass spectrometer for SRM analysis, e.g., TSQ Vantage triple-quadruple mass spectrometer (Thermo Scientific).
7. Nano-LC system, e.g., nano-Advance UHPLC system (Bruker Daltonics).
8. Analytical column, e.g., a self-made ESI column (*see* **Note 6**).
9. Trap column, e.g., L-column2 ODS (Chemicals Evaluation and Research Institute, Japan).
10. Software for iTRAQ-based MS analysis, e.g., Mascot (Matrix Science), which is used to identify protein and phosphorylation site and Proteome Discoverer 1.3 (Thermo Scientific), which is a platform to analyze qualitative and quantitative data of the identified proteins and phosphorylated peptides (*see* **Note 7**).
11. Software for SRM/MS analysis, e.g., Pinpoint (Thermo Scientific), software to obtain the peak areas (quantitative data of targeted peptides) from the raw data of SRM analysis as well as to develop the SRM methods.

12. 2 % acetonitrile, 0.1 % TFA.
13. 2 % acetonitrile, 0.1 % TFA, 25 μg/mL EDTA.
14. 1 pmol/μL BSA digest solution.

3 Methods

3.1 Grinding Frozen Tissues

1. Chill the stainless tissue pulverizer in liquid nitrogen. Place a piece of frozen tissue in the chilled device and pulverize the tissue by striking the device with a mallet several times.
2. Check how small the particles are that the tissue has been crushed into. Rearrange the particles with a chilled spoon and keep striking the device again until there are no large pieces left in the particles.
3. Transfer the grinded tissue into the chilled tube. Store at −80 °C.

3.2 Homogenizing Tissue

1. Place part of the grinded tissue into a microcentrifuge tube and weigh the amount required for analysis. In our case, at least 40 mg of tissue was used to obtain at least 3.17 mg protein (*see* **Note 8**). If the degree of contamination by blood is predicted to be high, wash the sample by adding an appropriate volume of PBS, spin down the pellet, and discard the supernatant.
2. Add cold lysis buffer, approximately 15 μL per 1 mg of tissue (e.g., for 40 mg of tissue, add 600 μL) (*see* **Note 9**). Suspend the grinded tissue lightly by pipetting.
3. Immediately homogenize the tissue by sonication, which is performed by placing the tubes containing the sample into an ice-water bath in the Bioruptor-UCD-250 and sonicating for 10 min (30 s on/30 s off) several times in the device with the amplitude set to 250 W (*see* **Note 10**).
4. Centrifuge the sample at 100,000 × *g* for 30 min at 4 °C. Collect the supernatant into a new tube. Place a small amount of the sample into another tube to determine the protein concentration. Store the remainder at −80 °C.
5. Dilute the sample for determining the protein concentration several times with TBS. The protein concentration is determined by a DC protein assay kit using BSA as the standard.

3.3 Protein Digestion

1. Add the protein extract to a new tube: 2 mg protein from individual tissue sample and 2 mg protein from standard mixture for iTRAQ analysis and 500 μg protein from individual tissue sample for SRM analysis. Dilute the protein with the lysis buffer to a concentration that is constant between all samples (*see* **Note 11**).

2. Reduce cysteine residues in the proteins in the homogenate with 10 mM DTT for 30 min at 37 °C.
3. Alkylate the residues with 50 mM IAA for 30 min at 37 °C in the dark.
4. Dilute the sample five times with PTS-A buffer.
5. Digest the proteins by 1:100 (w/w) LysC for 8 h at 37 °C.
6. Sequentially digest the sample by 1:100 (w/w) trypsin for 12 h at 37 °C.
7. Add an equal volume of ethyl acetate to the resulting peptide mixture. Acidify the sample by adding 1/200 volume of TFA (i.e., 0.5 % TFA) in order to transfer the detergents from the lysis buffer into the ethyl acetate layer while the peptides exist in the water layer. Mix the ethyl acetate layer and water layers well by vortexing the tube. Centrifuge the tube at 10,000 × *g* for 10 min at room temperature to separate both layers. Discard the upper ethyl acetate layer.
8. Dry the resulting peptide mixture using Speed Vac.
9. Store at −80 °C until starting the following enrichment of phosphopeptides.

3.4 Preparing Fe-IMAC Resin

1. Suspend Probond™ nickel-chelating resin in 20 % ethanol. Transfer the resin to empty spin columns (*see* **Note 12**). Centrifuge the resin at 150 × *g* for 2 min. Discard the flow-through.
2. Add 50 mM EDTA-Na solution (3 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min to release the nickel ions from the resin. Repeat this step twice more (*see* **Note 13**).
3. Add water (3 mL of water per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min. Discard the flow-through.
4. Add 1 % acetate solution (3 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min. Discard the flow-through. Repeat this step once more.
5. Add 100 mM $FeCl_3$ in 0.1 % acetic acid (2 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min to chelate iron ions to the resin. Repeat this step once more.
6. Add 1 % acetate solution (3 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min to wash the resin. Repeat this step twice more.
7. Add 60 % acetonitrile and 0.1 % TFA (3 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min. Discard the flow-through. Repeat this step once more.

8. Add 60 % acetonitrile and 0.1 % TFA to the resin with the bottom end of the column plugged and then suspend and transfer the resulting Fe-IMAC resins to a tube (*see* **Note 14**).

3.5 Enrichment of Phosphopeptides for the Large-Scale Analysis (iTRAQ Analysis)

1. Dissolve the peptide mixture prepared from the tissue sample in 60 % acetonitrile and 0.1 % TFA.
2. Add Fe-IMAC resin (1 mg of the resin for 2 mg of protein) into an empty spin column. Centrifuge at 150 × *g* for 2 min to discard the flow-through.
3. Add 60 % acetonitrile and 0.1 % TFA to the resin in the column. Centrifuge at 150 × *g* for 2 min. Discard the flow-through.
4. Load the peptide mixture in 60 % acetonitrile and 0.1 % TFA to the resin in the column. Centrifuge at 150 × *g* for 2 min. Discard the flow-through.
5. Add 60 % acetonitrile and 0.1 % TFA (3 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min to wash off the non-phosphopeptides. Repeat this step twice more.
6. Add 2 % acetonitrile and 0.1 % TFA (3 mL of the solution per 1 mL of the resin) to the resin in the column. Centrifuge at 150 × *g* for 2 min. Discard the flow-through.
7. Add 1 % phosphate solution (1 mL of the solution per 1 mL of the resin) to the resin in the column and centrifuge at 150 × *g* for 2 min. Collect the eluate into a tube. Repeat this step once more and then collect the second eluate into the same tube.
8. Desalt the eluate with a disposable solid-phase extraction (SPE) device such as Sep-Pak C18 or C18 Stage Tip.
9. Dry the sample using Speed Vac.
10. Store at −80 °C until iTRAQ labeling.

3.6 Labeling of Phosphopeptides by iTRAQ Reagents

1. Dissolve the sample enriched by IMAC in 30 μL of 1.0 M triethylammonium bicarbonate solution.
2. Return the iTRAQ reagent to room temperature and then add 70 μL of ethanol to each iTRAQ reagent vial.
3. Vortex each vial for 1 min to dissolve the iTRAQ reagent and then spin down.
4. Transfer one iTRAQ reagent to one sample tube (*see* **Note 15**). Vortex each tube to mix and then spin down.
5. Incubate the tubes at room temperature for 1 h.
6. Terminate the reaction by adding an equal volume of water. Vortex each tube to mix and then spin down.

7. Combine all of the iTRAQ-labeled samples in a new tube. Vortex the tube to mix and then spin down.
8. Dry the resulting mixture using Speed Vac.
9. Dissolve the sample in 2 % acetonitrile and 0.1 % TFA, and then check the pH. If the sample is not acidic, acidify by adding TFA. After acidifying, desalt the sample with a disposable SPE device such as C18 Stage Tip.
10. Dry the sample using Speed Vac.
11. Store at −80 °C until SCX fractionation.

3.7 Strong Cation-Exchange Chromatography

1. Dissolve the iTRAQ-labeled sample in SCX Buffer A (*see* **Note 16**).
2. Fractionate the sample using an HPLC system fit with an SCX column. Separate the sample using a gradient of SCX Buffers A and B and sequentially collect the eluted sample in microfuge tubes every 1 min (*see* **Note 17**).
3. Dry the fractions collected every 1 min using Speed Vac.
4. Decrease the number of the fractions for the subsequent MS analysis by combining the fractions collected every 1 min based on the peak intensity on the HPLC chromatogram (*see* **Note 18** and Fig. 2).
5. Desalt the combined fractions with a disposable SPE device such as C18 Stage Tip.
6. Elute the sample into a sample vial for MS analysis and then dry it using Speed Vac.
7. Store at −80 °C until MS analysis.

3.8 Comprehensive Analysis by LC-MS/MS

1. Add 10 μL of 2 % acetonitrile and 0.1 % TFA to each sample vial.
2. Vortex each vial for 1 min to dissolve the fractionated peptides and then spin down.
3. Set the operating parameters of the mass spectrometer for iTRAQ analysis (*see* **Note 19**).
4. Analyze each sample by LC-MS/MS (*see* **Note 20**).
5. Apply the acquired raw file to the search software such as MASCOT to identify and quantify the phosphopeptides (*see* **Note 21**).
6. Select the phosphopeptide that has to be validated in the subsequent SRM analysis by quantitatively comparing the identified phosphopeptides.

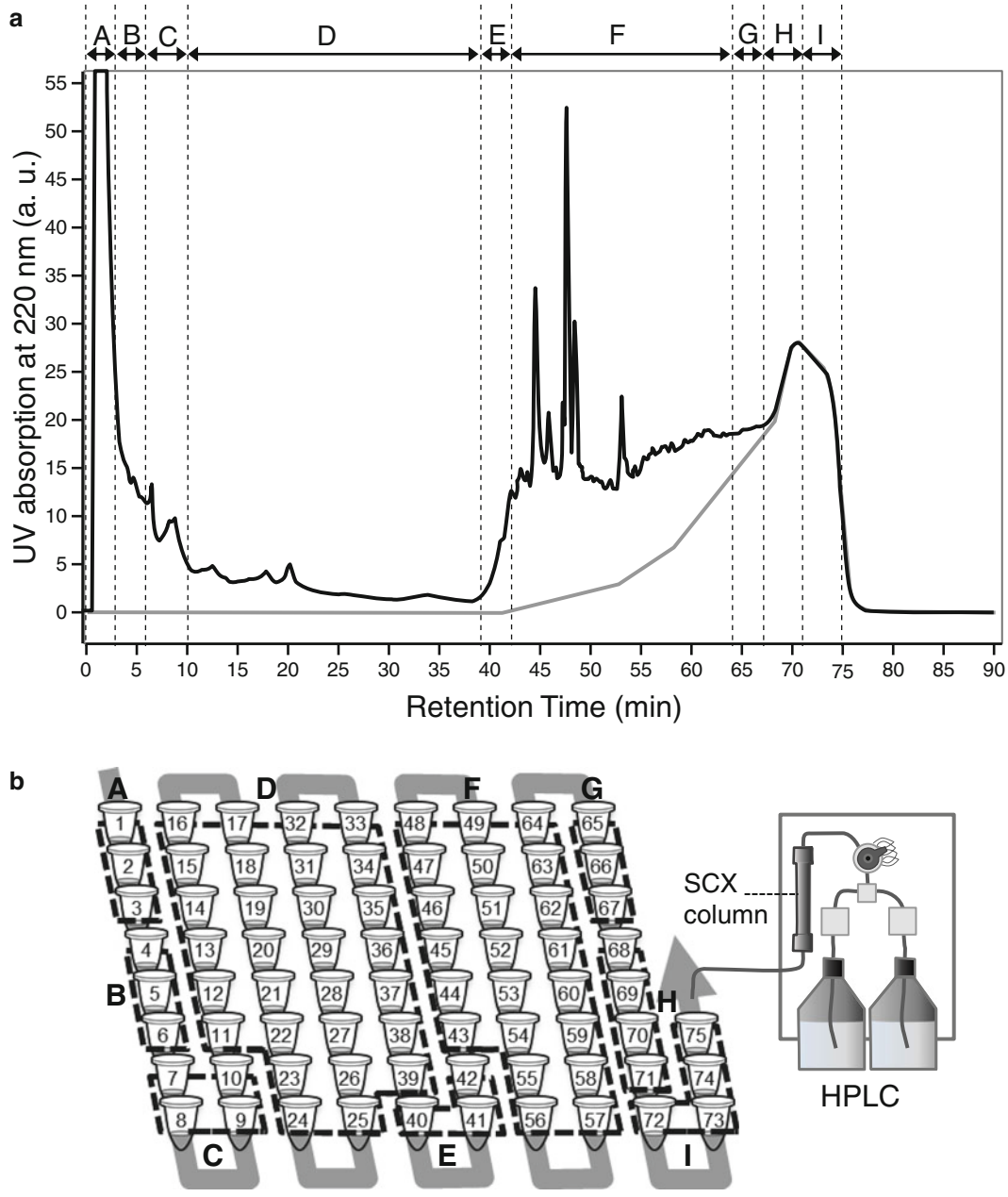


Fig. 2 A SCX chromatography from our study is shown in (**a**). The black trace is the result of iTRAQ-labeled phosphopeptides, which was prepared as indicated in Fig. 1. The gray trace is the baseline. The eluate of the chromatography was fractionated into 75 fractions every 1 min (**b**). The fractions were separated into nine groups (Group-A–I) in a manner dependent on the intensity in the chromatogram as shown in (**a**) and (**b**). The eight groups other than the group-F have low intensities and so all fractions of their groups are each combined into a single fraction

3.9 Sample Preparation for SRM Analysis (for up to 100 μg of Proteins (See Note 22)

1. Prepare a homogenate by processing tissue samples according to the "Homogenizing Tissue" section.
2. Digest the homogenate according to "Protein Digestion" section, **steps 1** and **6**, and prepare the peptide mixture from which the detergent has not yet been removed.
3. Add all SI peptides to each sample (*see* **Notes 23** and **24**).
4. Extract the peptides from the sample according to **steps 7–9** in "Protein Digestion" section.
5. Desalt the resulting sample with a disposable SPE device such as C18 Stage Tip.
6. Prepare the IMAC-C18 Stage Tip by packing 2 layers of C18 resin at the end of a 200 μL pipet tip and then loading 50 μL of Fe-IMAC resin on the C18 resin. Stick the IMAC-C18 Stage Tip to the microcentrifuge tube with a hole made in the center of the lid and remove the solution by centrifugation at 800 × *g* for 2 min.
7. Load the desalted sample in 60 % acetonitrile and 0.1 % TFA to the IMAC-C18 Stage Tip and then centrifuge at 600 × *g* for 5 min.
8. Add 200 μL of 60 % acetonitrile and 0.1 % TFA to the IMAC-C18 Stage Tip and then centrifuge at 800 × *g* for 4 min to wash non-phosphopeptides off the IMAC resin. Repeat this step twice more.
9. Add 200 μL of 0.1 % TFA to the IMAC-C18 Stage Tip and then centrifuge at 800 × *g* for 4 min to equilibrate the C18 resin under the IMAC resin.
10. Add 100 μL of 1 % phosphate to the IMAC-C18 Stage Tip and then centrifuge at 800 × *g* for 2 min to release the phosphopeptides from the IMAC resin and then bind them to the C18 resin. Repeat this step once more.
11. Add 200 μL of 0.1 % TFA to the IMAC-C18 Stage Tip and then centrifuge at 2,300 × *g* for 2 min to wash the C18 resin.
12. Elute the phosphopeptides bound to the C18 resin to a sample tube using 60 μL of 60 % acetonitrile and 0.1 % TFA.
13. Dry the sample using Speed Vac.
14. Store it at −80 °C until MS analysis.

3.10 Targeted Analysis by LC-SRM/MS

1. Analyze a mixture of the stable isotope-labeled peptide (SI peptides), which has the same sequence as the phosphopeptide selected in the former comprehensive analysis, by LC-MS/MS.
2. Create a primary method for the subsequent SRM analysis by analyzing the acquired MS data (*see* **Note 25**) and selecting the precursor ions of each target observed with a strong signal intensity (doubly, triply, or higher charged ions) and the

product ions generated from the precursor ion with a strong signal intensity (*see* **Note 26**).

3. Optimize the parameters (*m/z* of product ions and CE) of the SRM method by LC-SRM/MS (*see* **Note 27**).
4. Add 10 μL of 2 % acetonitrile, 0.1 % TFA, and 25 μg/mL EDTA to each sample.
5. Vortex each vial for 1 min to dissolve the fractionated peptides and then spin down.
6. Set the optimized SRM method and other operating parameters for the SRM analysis (*see* **Note 28**).
7. Analyze each sample by LC-SRM/MS (*see* **Note 29**).
8. Apply the acquired raw data to the software for quantification (*see* **Note 30**). Quantitatively compare the target peptides between the samples by calculating the peak area in a chromatogram of each SRM transition and then normalizing the values of the endogenous targeted peptides to those of the corresponding SI peptides.

4 Notes

1. The procedures used for homogenizing tissue samples and enzymatic digestion are based on phase transfer surfactant (PTS)-aided trypsin digestion as described in a previous study [17].
2. Add 1 mL of water to a bottle containing 1 g of lyophilized LysC.
3. Add 100 μL of 10 mM HCl to a bottle containing 100 μg of lyophilized trypsin.
4. Peptide mixtures are desalted using C18 Stage Tip or another solid-phase extraction (SPE) device such as Oasis HLB. Desalting by C18 Stage Tip is performed as described in a previous study [18]. Briefly, a small 47 mm Empore™ C18 disk is stamped out using a blunt-ended syringe needle (16 G), and then the layers are placed in a 200 μL pipet tip by pushing them from the top of the tip using a plunger. Methanol (for swelling), 80 % acetonitrile, 0.1 % TFA (for washing), and 2 % acetonitrile and 0.1 % TFA (for equilibrating) are passed through by centrifugation the C18 resin in this order. After the sample is passed through to absorb the peptides to the C18 resin, 2 % acetonitrile, and 0.1 % TFA (for washing) is passed through. Elution of the peptide mixture is performed by 60 % acetonitrile and 0.1 % TFA. Loading capacity is 20 μg per layer of C18 resin. The volume of all solutions is 20 μL per layer of C18 resin. When another SPE device is used, desalting is performed according to the instructions of each manufacturer.

5. We mostly replace lysine or arginine at the C-terminal of target peptides with isotope-labeled lysine (13C6, 15 N2) or arginine (13C6, 15 N4) in order to make y-ions heavier. When the amino acid at the C-terminal is not lysine or arginine (e.g., the C-terminal of a protein), we replace the other amino acids (e.g., alanine) at or near the C-terminal with the other isotope-labeled one (e.g., Alanine-13C3–15 N1).
6. We make an ESI column by packing C18 particles by a capillary column packer into a glass capillary needle (200 mm length × 100 μm for the inner diameter) which is made by laser puller.
7. By using Proteome Discoverer 1.3, we obtain the list in which there are the identified phosphopeptides associated with quantitative data (quantitative values obtained from iTRAQ-reporter ions) and qualitative data (Mascot ion score and probability of phosphorylation sites). We export the list into an excel file and follow by editing it (filtering the phosphopeptides by score and merging the results of multiple analysis).
8. In our study, we need 2 mg protein +0.67 mg protein (to make standard mixture) for iTRAQ analysis and 0.5 mg protein for SRM analysis prepared from each sample. To obtain enough amount of protein, we use more than 40 mg of tissue if possible.
9. By adding the buffer to samples at this ratio, we can generally obtain a solution containing 5–15 mg of proteins per mL.
10. After several rounds of sonication, we examine the contents in the tubes in order to check the residual pieces of the tissue. If the tissues are completely dissolved, we stop the sonication. If not, a few rounds of sonication are additionally performed until the size of residual tissues remains the same. At this stage, we consider the proteins to be sufficiently extracted from the tissue and stop the sonication.
11. We use 2 mg of protein from individual tissue sample and 2 mg of protein from standard mixture for iTRAQ analysis and 500 μg protein from individual tissue sample for SRM analysis. The reasons for the amounts of the protein used are as follows. Maximum amount of peptides to load on our LC-MS system is 2–3 μg considering the separation ability of our analytical column with an inner diameter of 100 μm and the robustness of the LC-MS systems. As shown in Figs. 1 and 2, less than 80 μg phosphopeptide is estimated to be obtained for iTRAQ analysis after phosphopeptide enrichment and less than 3 μg phosphopeptide in each combined fraction after SCX chromatography because the amount of phosphoprotein is estimated to be 1 % of the total protein (also accounting for sample loss during our procedures). Similarly, in case of SRM analysis, we estimate less than 3 μg phosphopeptide to be obtained after phosphopeptide enrichment.

12. The ProBond resin is initially provided as 50 % slurry in 20 % ethanol. We use 1 mL of the resin (or 2 mL of the suspension) for up to 2 mg of proteins.
13. When nickel ions are released from the resin by EDTA and washed away, the color of the resins turns from blue to white. If the nickel ions are not released sufficiently based on the color of the resins, repeat this step once more.
14. We store the Fe-IMAC resin as 50 % slurry at 4 °C and use it within 1 week.
15. When we quantitatively compare more than 4 samples by iTRAQ analysis, we use iTRAQ 115, 116, and 117 to label individual samples and iTRAQ 114 as the reference sample. This reference sample is the mixture of an aliquot of all samples and is used as the standard in all iTRAQ experiments.
16. The volume of SCX buffer A needed to dissolve the sample depends on the HPLC systems used. We dissolve the sample in 110 μL of SCX buffer A according to the maximum volume (100 μL) of the autosampler in our HPLC system. 100 μL of the sample is loaded onto the HPLC equipment. The remainder (10 μL) is used to assess iTRAQ labeling by MS analysis.
17. We use a flow rate of 200 μL/min and four-step linear gradient for the separation, as follows: 0 % B for 30 min, 0–10 % B in 15 min, 10–25 % B in 10 min, 25–40 % B in 5 min, 40–100 % B in 5 min, and 100 % B for 10 min.
18. We combine the 75 fractions collected every 1 min into 30 fractions. The flow-through fraction is not combined in case polymer-like contaminants are found in it. If the fractions collected every 1 min are collected at the time when the peak intensities are lower in the HPLC chromatogram, a larger number of samples is combined. If the fractions are collected when the peak intensities are higher, we combine a few fractions or use it for MS analysis as a single fraction (*see* Fig. 2).
19. When we perform iTRAQ analysis using the LTQ-Orbitrap XL or Velos mass spectrometer, the operating parameters are set as follows: full MS scans are performed in the orbitrap mass analyzer (scan range 350–1500 m/z, with 30 K FWHM resolution at 400 m/z). The three (LTQ XL) or five (LTQ Velos) most intense precursor ions are selected for the MS/MS scans. MS/MS scans are performed using collision-induced dissociation (CID) and higher energy collision-induced dissociation (HCD, 7500 FWHM resolution at 400 m/z) for each precursor ion. Collision energy is set to 35 % for CID and 50 % for HCD. A dynamic exclusion option is implemented with a repeat count of 1 and exclusion duration of 60 s. The values of automated gain control (AGC) are set to 5.00e+05 for full MS, 1.00e+04 for CID MS/MS, and 5.00e+04 for HCD MS/MS.

20. We analyze the fractionated peptides using an LTQ-Orbitrap XL or Velos mass spectrometer equipped with a nano HPLC system and HTC-PAL autosampler. The analytical column is self-made by packing C18 particles (L-column2 ODS, 3 μm) into a self-pulled needle (200 mm length × 100 μm for the inner diameter). The mobile phases consist of buffers A (0.1 % formic acid, 2 % acetonitrile) and B (0.1 % formic acid, 90 % acetonitrile). Samples are loaded onto the trap column. The nano LC gradient is delivered at 500 nL/min and consists of a linear gradient of Buffer B developed from 5 to 30 % B in 135 min. A spray voltage of 2000 V is applied.
21. To identify the phosphopeptides, the CID and HCD raw spectra are extracted and searched separately against the forward and reverse-decoy human IPI database (version 3.68) using Proteome Discoverer 1.3 and Mascot v2.2. The precursor mass tolerance is set to 3 ppm and a fragment ion mass tolerance is set to 0.6 Da for CID and 0.01 Da for HCD. The search parameters allow for one missed cleavage for trypsin, fixed modifications (carbamidomethylation at cysteine and iTRAQ labeling at lysine and the N-terminal residue), and variable modifications (oxidation at methionine, iTRAQ labeling at tyrosine, and phosphorylation at serine, threonine, and tyrosine). The score threshold for peptide identification is set at 1 % false-discovery rates (FDR). Peptides identified at a threshold with 5 % FDR are also accepted in cases in which the peptide with the same sequence is identified at a threshold with 1 % FDR in any three other iTRAQ experiments. To quantify the phosphopeptides, we obtain the iTRAQ quantitation values automatically calculated from the intensity of the iTRAQ reporter ions in the HCD scans using the Proteome Discoverer 1.3. Quantitation of peptides identified from CID scans is performed using the reporter ion information extracted from the HCD spectra of the same precursor peptide. In cases in which peptides with the same sequence are identified repeatedly from different precursor peptides in the same iTRAQ experiment, the median of their quantitation values is calculated. The iTRAQ quantitation values of individual samples (iTRAQ 115, 116, and 117) are normalized with the values of the reference sample (iTRAQ 114) in each iTRAQ experiment for comparisons among all of iTRAQ experiments.
22. We use tissue samples containing 500 μg of proteins for SRM-based validation of the results of our large-scale phosphoproteome analysis. Each homogenate containing 500 μg of proteins is equally divided into five tubes, and processing is then performed according to the protocol.
23. In the case of crude SI peptides, the purities are very different between the products especially for phosphopeptides. In addition, the ionization efficiency depends on peptide

sequences. As a result, the signal intensities of SI peptides can be very different. To maintain the robustness of the experimental system, the signal intensity of each SI peptide should be checked by LC-MS/MS before mixing and then the amount of each SI peptide added should be modified based on the signal intensity.

24. When we perform serial dilutions of the SI peptides for the addition of small amounts of peptides, we perform the dilution with a 1 pmol/μL BSA digest solution to prevent adsorption.
25. The SRM method consists of SRM transitions, which mean pairs of m/z of the precursor/product ions, the collision energies (CEs), and retention time of the nano HPLC.
26. In our case, the SI peptide mixture is analyzed by LC-MS/MS using LTQ-Orbitrap XL (CID mode) and the msf file is generated using Proteome Discoverer and Mascot. The msf file is opened with Pinpoint software (version 2.3.0, Thermo Scientific) and the list of MS/MS fragment ions derived from SI-peptides is generated. A total of multiple product ions (4–10 product ions) are selected for the SRM transitions of each target peptide based on the following criteria: y-ion series, strong ion intensity, at least 2 amino acids in length, and no neutral loss fragment.
27. At first, we optimize collision energy (CE) for every SRM transition around the theoretical value calculated according to the formulas; $\mathrm{CE} = 0.044 \times m/z + 5.5$ for doubly charged precursor ions, and $\mathrm{CE} = 0.051 \times m/z + 0.55$ for triply charged precursor ions. In cases in which the theoretical value is over 35 eV, the value is set to 35 eV. After this optimization, the 4 most intense transitions are selected for each target peptide.
28. In addition to the SRM method (SRM transitions, CE, and the retention time for the each target peptide), the parameters of the instrument are set as follows; a scan width of 0.002 m/z, Q1 resolution of 0.7 FWHM, cycle time of 1 s, and gas pressure of 1.8 mTorr. Data are acquired in the time-scheduled SRM mode (retention time window: 8 min).
29. We use the TSQ-Vantage triple quadruple mass spectrometer equipped with the LC system described above. The nanoLC gradient is delivered at 300 nL/min and consists of a linear gradient of mobile phase B developed from 5 to 23 % B in 45 min. A spray voltage of 1800 V is applied.
30. We use Pinpoint, software for the analysis of SRM data. SRM transitions with more than 1×10^3 ion intensity at the peak are used for quantitation. We check that the ratios among the peak areas of individual SRM transitions for each targeted phosphopeptide are comparable to those of the corresponding SI peptide.

References

1. Neville DC, Rozanas CR, Price EM et al (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. Protein Sci 6: 2436–2445
2. Villen J, Beausoleil SA, Gerber SA et al (2007) Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci U S A 104: 1488–1493
3. Ballif BA, Villen J, Beausoleil SA et al (2004) Phosphoproteomic analysis of the developing mouse brain. Mol Cell Proteomics 3: 1093–1101
4. Beausoleil SA, Jedrychowski M, Schwartz D et al (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101:12130–12135
5. Ficarro SB, McCleland ML, Stukenberg PT et al (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol 20: 301–305
6. Moser K, White FM (2006) Phosphoproteomic analysis of rat liver by high capacity IMAC and LC–MS/MS. J Proteome Res 5:98–104
7. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd et al (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 316:1160–1166
8. Olsen JV, Blagoev B, Gnad F et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127: 635–648
9. Trinidad JC, Thalhammer A, Specht CG et al (2008) Quantitative analysis of synaptic phosphorylation and protein expression. Mol Cell Proteomics 7:684–696
10. Nguyen V, Cao L, Lin JT et al (2009) A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation. Mol Cell Proteomics 8:2418–2431
11. Ong SE, Blagoev B, Kratchmarova I et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376–386
12. Ross PL, Huang YN, Marchese JN et al (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154–1169
13. Muraoka S, Kume H, Watanabe S et al (2012) Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples. J Proteome Res 11:4201–4210
14. Kume H, Muraoka S, Kuga T et al (2014) Discovery of Colorectal Cancer Biomarker Candidates by Membrane Proteomic Analysis and Subsequent Verification using Selected Reaction Monitoring (SRM) and Tissue Microarray (TMA) Analysis. Mol Cell Proteomics 13:1471–1484
15. Narumi R, Murakami T, Kuga T et al (2012) A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. J Proteome Res 11:5311–5322
16. Lange V, Malmstrom JA, Didion J et al (2008) Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. Mol Cell Proteomics 7:1489–1500
17. Masuda T, Tomita M, Ishihama Y (2008) Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J Proteome Res 7:731–740
18. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2: 1896–1906

Part III

Phosphopeptide Enrichment Strategies

Chapter 7

Enrichment Strategies in Phosphoproteomics

Alexander Leitner

Abstract

The comprehensive study of the phosphoproteome is heavily dependent on appropriate enrichment strategies that are most often, but not exclusively, carried out on the peptide level. In this chapter, I give an overview of the most widely used techniques. In addition to dedicated antibodies, phosphopeptides are enriched by their selective interaction with metals in the form of chelated metal ions or metal oxides. The negative charge of the phosphate group is also exploited in a variety of chromatographic fractionation methods that include different types of ion exchange chromatography, hydrophilic interaction chromatography (HILIC), and electrostatic repulsion HILIC (ERLIC) chromatography. Selected examples from the literature will demonstrate how a combination of these techniques with current high-performance mass spectrometry enables the identification of thousands of phosphorylation sites from various sample types.

Key words Phosphopeptide enrichment, Fractionation, Sample preparation, Mass spectrometry

1 Introduction

Recently, two high-profile reports from two independent consortia have reported the completion of drafts of the human proteome [1, 2]. Regardless of whether one agrees that this milestone has actually been achieved or not, the enormous methodological and technological advances in recent years have made it possible to profile many proteomes to a considerable depth. Although similar advances have been made in the area of profiling posttranslational modifications (PTM), it is evident that the community is still quite far from reporting the comprehensive characterization of any PTM in a biological system. However, protein phosphorylation is quite likely the most widely studied of such modifications, and its huge importance for a variety of biological processes has fueled the development of an extensive suite of analytical methodologies.

The current state of phosphoproteomics allows the identification of serine, threonine and tyrosine phosphorylation to an unprecedented depth [3–6]. This, in turn, has allowed new insights into the organization of signaling transduction pathways and other

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_7,

processes mediated by phosphorylation [7–9]. In addition, other types of phosphorylation, for example on histidine or arginine residues, are becoming more accessible due to newly generated antibodies and analytical workflows (see below). In combination with the increased sensitivity and sequencing speeds of mass spectrometers, the analysis of samples of limited availability, such as tissue biopsies, is within reach. This means that aberrant signaling pathways cannot only be studied on a system-wide level in samples derived from cell culture, but eventually directly from patient-derived tissue specimens.

To achieve this, the phosphoproteomic workflow needs to deal with some well-known challenges of analyzing phosphopeptides by mass spectrometry:

- Proteins with important roles in signaling pathways are frequently less abundant.
- Phosphorylation is often substoichiometric, and modified peptides have to be detected in an excess number of unphosphorylated peptides.
- Phosphopeptides may have altered physicochemical properties (such as hydrophilicity or proton affinity) that are relevant for their identification by LC-MS, although these effects cannot be generalized.
- Phosphopeptides may show unfavorable fragmentation behavior in collision-induced dissociation (CID), the most common tandem mass spectrometry technique.

Efficient sample preparation protocols can help to overcome many, but not all of these challenges. Here, I will discuss the most relevant enrichment and fractionation methods that are currently used in large scale phosphoproteomics projects (*see* Subheading 3). Detailed protocols for several of these methods are presented in the following chapters of this volume, and they are cross-referenced. To put these experimental strategies into a better context, I will also give a brief overview of mass spectrometric methodology related to phosphorylation profiling in Subheading 2. Again, a number of these methods are covered in detailed protocols later in this book. Finally, I will illustrate what can be achieved by the intelligent combination of optimized enrichment methods and powerful downstream MS analysis with selected applications taken from the recent literature that are highlighted in Subheading 4.

2 Large-Scale Analysis of Protein Phosphorylation by Mass Spectrometry

Although the mass spectrometric analysis of phosphopeptides shares many conceptual similarities with that of unmodified peptides, particular properties of phosphopeptides make it more challenging. The expanding capabilities of mass spectrometers have led

to the introduction of new acquisition schemes dedicated to phosphopeptide analysis. As mentioned above, many difficulties with phosphopeptide analysis result from their low abundance, or the low abundance of the phosphoproteins themselves, which can be dealt with by enrichment strategies described in the following section.

The remaining challenges are mainly related to the fragmentation properties of phosphopeptides in tandem mass spectrometry experiments. CID remains the predominant fragmentation method in mass spectrometry, whereby the carbon-nitrogen bond of the peptide bond is cleaved upon collision with gas molecules. Unfortunately, peptides phosphorylated on serine or threonine (but not tyrosine) residues are known to undergo a preferential loss of the phosphate group upon CID, resulting in a net loss of approximately 98 Da, corresponding to H_3PO_4. This effect is particularly dominant in ion trap CID (itCID), where the low-energy resonance excitation process favors energetically labile bonds [10, 11]. In extreme cases, the predominant neutral loss results in strong suppression of the desired fragmentation of the peptide backbone that leads to the formation of series of b and y ions, making sequence assignment impossible. To overcome this limitation, a number of modified or alternative fragmentation methods have been proposed. For example, a consecutive stage of fragmentation (MS^3) in instruments equipped with an ion trap analyzer can target the neutral loss product, although this complicates data analysis and compromises scan speed and sensitivity. Multi-stage activation (MSA) is a related concept, but avoids the additional isolation step by consecutive activation of the precursor and putative neutral loss products. MS^3 and MSA strategies are discussed in more details elsewhere [5, 11]. Fortunately, the increased dynamic range of modern (linear) ion trap analyzers has made it increasingly possible to generate good-quality CID-MS/MS data even in the presence of dominant neutral loss signals.

In recent years, instruments that employ collision cell CID (ccCID) have again experienced increased popularity. This is due to the availability of faster quadrupole-time-of-flight (Q-TOF) instruments and the introduction of ccCID on Orbitrap-type instruments, where it is called higher energy collisional dissociation (HCD). Conceptually, ccCID (or beam-type CID, as it is also called) is very similar to trap CID, leading to cleavage of the amide bond to yield b and y ions, although it differs in the time scale of energy deposition and energy range [10]. Practically, ccCID spectra of phosphopeptide show less dominant neutral loss and a stronger preference for y ion series, due to consecutive fragmentation of the more labile b ions.

Another alternative is the use of the electron-based dissociation techniques, electron capture dissociation (ECD), and electron transfer dissociation (ETD). Both methods result in cleavage of the

backbone N-C_α bond, either after the capture of a free electron (in ECD) or transfer of an electron from a donor molecule (in ETD). While ECD is practically restricted to expensive and not widely available ion cyclotron resonance mass spectrometers, ETD is commercially implemented on both linear ion trap-Orbitrap and Q-TOF hybrid instruments. Because of the fundamental differences in the fragmentation mechanisms compared to CID, ECD and ETD result in the formation of c and z ion series instead of b and y ions. Furthermore, the activation process does not lead to a preferential neutral loss from phosphorylated side chains. Therefore, ECD/ETD spectra typically result in more bond cleavages in phosphopeptides than the corresponding CID spectra. The main practical drawbacks of electron-based fragmentation of phosphopeptides, however, are the low fragmentation efficiency, resulting in lower sensitivity, and the relatively long activation times (especially compared to ccCID in modern instruments [5]).

The second level of complexity when dealing with the mass spectrometric identification of phosphopeptides is related to the exact localization of the phosphorylation site(s) in the peptide [12]. While all fragment ions can contribute to the assignment of the main peptide sequence, pinpointing the precise location of the phosphorylation site requires high-quality MS/MS data. This is because frequently there are many potential phosphorylation sites within the peptide sequence, and in the "worst" case (two adjacent residues being potential modification sites), cleavage of a particular bond (namely between these residues) is required to exactly localize the modification.

The challenges of MS and MS/MS analysis of phosphopeptides discussed above mean that any phosphoproteomic workflow will benefit from samples that are enriched in phosphopeptides. This will increase the relative abundance of the actual targets of interest in the sample and decrease the overall sample complexity, thereby allowing for more flexibility in the mass spectrometric strategy.

3 Overview of Enrichment and Fractionation Concepts in Phosphoproteomics

3.1 Overview

Broadly speaking, dedicated phosphoproteomic sample preparation methods may be classified according to several criteria (*see* Fig. 1). First, they can be performed at the protein or at the peptide level. However, performing generic, sequence-independent enrichment methods at the protein level is not very effective—more than 50 % of all human proteins have already been found to be phosphorylated at some point [2]. Thus, a method that would indiscriminately enrich all phosphorylated proteins from a sample would not result in a substantial reduction of sample complexity. Therefore, protein-level enrichment is ideally performed using motif-specific antibodies directed at a specific phosphorylation site.

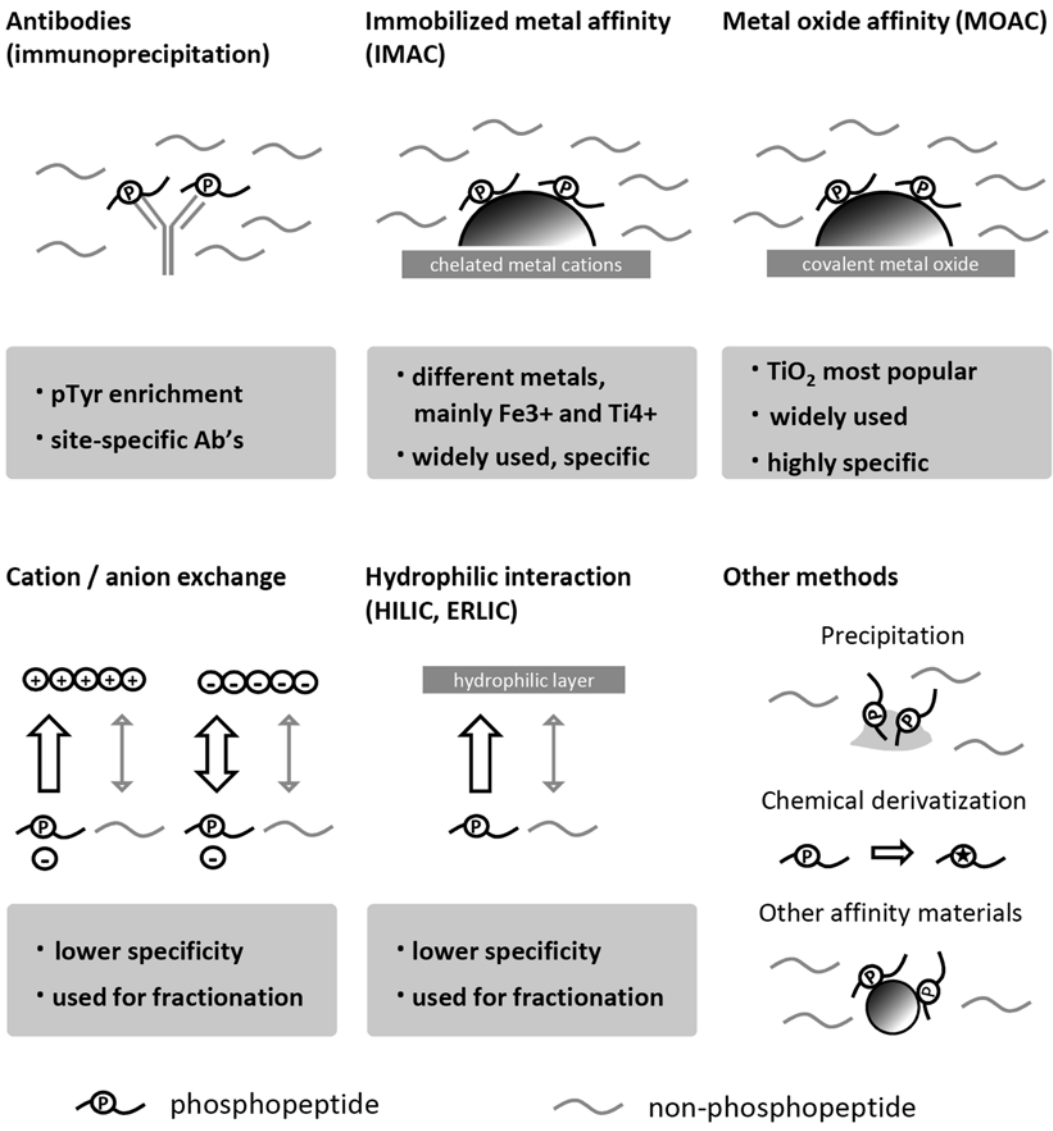


Fig. 1 Overview of analytical strategies in phosphoproteomics

To achieve a more comprehensive coverage, targeting the (phospho)peptide level is a more promising and practical approach. For this purpose, a number of different methods are available that exploit different recognition concepts (*see* Fig. 1). Some of them are more specific and used in a stepwise procedure (that is, binding and elution), and are summarized under the term *enrichment methods* in the following. Other methods cannot discriminate as efficiently between phosphopeptides and unphosphorylated peptides, and are commonly performed in the form of column liquid chromatography, whereby multiple fractions are collected for further

analysis. These methods are therefore classified as *fractionation methods*. In comprehensive experimental designs, fractionation and enrichment methods are often combined (*see* Subheading 3.8).

The most common enrichment methods introduced in the following involve immunoprecipitation using specific antibodies (*see* Subheading 3.2) and interaction with chelated metal ions (immobilized metal affinity chromatography (IMAC), *see* Subheading 3.3) or covalent metal oxides (metal oxide affinity chromatography (MOAC), *see* Subheading 3.4). Chromatographic fractionation methods may involve ion-exchange chromatography (*see* Subheading 3.5) and hydrophilic interaction chromatography (HILIC) or its variant ERLIC (electrostatic repulsion HILIC), both of which will be discussed in Subheading 3.6. Finally, other methods such as precipitation methods or methods involving the chemical derivatization of the phosphate group are summarized in Subheading 3.7, although they are not widely used for the proteome-wide study of phosphorylation.

3.2 Antibody-Based Methods

The generation of high-quality antibodies that recognize phosphorylated residues in a sequence-independent context has proven challenging, with phosphotyrosine (pTyr) as a notable example. pTyr-containing peptides can be enriched with high specificity, which is necessary to deal with the low abundance of such peptides that leads to their underrepresentation even when conventional phospho-specific enrichment methods are being employed [13, 14]. The practical limitation of pTyr profiling is the required amount of starting material—at least several mg of protein are needed to achieve adequate coverage, and some large-scale studies have used double-digit mg levels to obtain sufficient peptide amounts for mass spectrometric detection. An alternative recognition concept directed towards pTyr residues involves the use of SH2 domains.

Dedicated antibodies are also promising for the analysis of His phosphorylation, although the recently developed, relatively specific anti-pHis antibodies [15] still reported substantial cross-reactivity with pTyr.

3.3 Immobilized Metal Affinity Chromatography

IMAC is the oldest affinity chromatography technique that takes advantage of the interaction between the phosphate group and metals. In this case the metal is present in the form of a chelated metal ion that is non-covalently attached to a matrix via a chelating group such as nitrilotriacetate or iminodiacetate. Originally, the material was (and still is) used for purification of His-tagged proteins, and the application to the purification of phosphoproteins and peptides emerged only later [16, 17]. Currently, Fe^{3+}-IMAC is the most commonly used variant, and a large variety of IMAC materials are commercially available. The most significant drawback of conventional IMAC materials is the relatively low enrichment specificity (in comparison to metal oxides), because non-phosphorylated peptides with multiply acidic residues tend to show

strong nonspecific binding. Dedicated optimization of the enrichment protocol can improve the specificity of the approach, if desired. Alternatively, new-generation mass spectrometers with high sequencing speed are better able to deal with a larger fraction of non-phosphorylated peptides. Due to its ease of use and the availability of robust protocols, these methods continue to be widely applied alone or as part of multidimensional strategies [3–6]. A respective protocol is given in Chapter 8 by Thingholm and Larsen.

Recent notable innovations related to IMAC materials use different matrices for chelation. The Ti^{4+}-IMAC material introduced by Zou and co-workers [18, 19] uses a phosphonate moiety for immobilization, and Ti^{4+} is used as the metal cation. This material has shown impressive performance in recent large-scale applications (*see* Subheading 4). Tao and co-workers have introduced a dendrimer IMAC material [20, 21]. Dendrimers are soluble high-molecular-weight polymers that allow for improved binding kinetics and straightforward isolation via ultrafiltration. The latter study [21] also demonstrated the complementarity of IMAC materials with different chelated metals, in this case Ti and Zr. A similar approach, but using conventional Ga^{3+}- and Fe^{3+}-IMAC matrices, was followed by Chen and co-workers [22]. They could show that multiply phosphorylated peptides and phosphopeptides carrying additional acidic residues could be preferentially enriched on the weaker affinity Ga^{3+} material, while other singly phosphorylated peptides could be subsequently captured on the Fe^{3+} beads.

3.4 Metal Oxide Affinity Chromatography

MOAC is based on the interaction of the negatively charged phosphate with metal in the form of covalent oxides. The interaction mechanisms are assumed to consist of both ion exchange mechanisms and Lewis acid/base interactions, but have not been thoroughly characterized for large biomolecules such as peptides. Historically, titanium dioxide (titania, TiO_2) has been the first metal oxide to be used for the purpose of phosphopeptide enrichment [23, 24], although in recent years many other oxides have been found to exhibit specific affinity towards phosphopeptides to some degree. These include oxides of zirconium, aluminum, iron, or tin, just to name a few (for a detailed overview, see [25]). Due to their commercial availability, TiO_2 and ZrO_2 are predominantly used, with the majority of published protocols relying on titania (see also protocols by Thingholm and Larsen and Batth et al. in Chapters 9 and 12). Many nanoscale metal oxide materials have also been prepared and applied to more or less complex mixtures (*see* review Chapter 13 by Batalha et al.).

A generic MOAC protocol consists of the binding of peptides under strongly acidic conditions, followed by washing steps to reduce nonspecifically bound material, and an elution step under alkaline conditions. Over the years, a huge number of variations of this protocol have been introduced. Most of them use different binding and washing conditions.

Although MOAC is generally assumed to be more specific towards the binding of phosphopeptides than IMAC, nonspecific binding can be further optimized by controlling the loading step. For example, a pH of 2.7 or lower is commonly used to protonate the side chains of aspartic and glutamic acid residues, thereby reducing their tendency for ionic interactions. Further increase of specificity is possible by working either at very low pH using high concentration of strong acids such as 5 % trifluoroacetic acid, or by using hydroxy acids that compete with non-phosphopeptides for low affinity binding sites. These additives include 2,5-dihydroxybenzoic acid [24], lactic acid [26], and glycerol [27], among others.

Elution from the metal oxides is usually performed with a diluted ammonium hydroxide solution, although some groups have reported improved recoveries of very strongly bound phosphopeptides by adding phosphate salts or performing additional elution using organic amines [28].

The combination of binding and elution conditions, but also the specific type of material [29] and the peptide-to-bead ratio [30] have all been found to contribute to the performance of the method in terms of specificity and bias towards mono- and poly-phosphorylated peptides. A modified TiO_2 enrichment protocol to study arginine phosphorylation has recently been introduced [31], whereby peptide binding is carried out at a higher pH to avoid degradation of this acid-labile modification.

3.5 Ion-Exchange Chromatography

Both strong cation-exchange (SCX, [32, 33]) and strong anion-exchange (SAX, [34]) chromatography have been used to partially separate phosphopeptides from their non-phosphorylated counterparts, with SCX being more frequently used. The underlying separation principle is simple: In SCX, the strongly acidic phosphopeptides interact poorly with the anionic stationary phase and elute early from the column. In SAX, the opposite mechanisms is at play and phosphopeptides are retained more strongly on the cationic stationary phase.

In both cases, the separation is based on the different solution charge state of tryptic peptides and phosphopeptides. If Asp and Glu side chains are protonated at low pH, a typical tryptic, unphosphorylated peptide is expected to have a charge state of at least +2 (C-terminal Arg or Lys and N-terminal amine group). This charge state is reduced to +1 by addition of a negatively charged phosphate group. Addition of further phosphates in multiply phosphorylated peptides may extend the charge separation. However, there are also several cases where this ideal charge state difference does not hold. For example, N-terminally blocked (e.g., *N*-acetylated) peptides or peptides carrying other acidic modifications also have lower charge states, while missed cleavage sites may cause higher charge states even for phosphopeptides. In practice, this results in a reduced discrimination of ion exchange-based chromatographic separation. This effect and the overall relatively

low separation efficiency of ion-exchange chromatography make this method a better choice for integration in a multidimensional method for phosphoproteome analysis (*see* Subheading 3.8).

3.6 Hydrophilic Interaction Chromatography and Electrostatic Repulsion HILIC

Recently, variants of hydrophilic interaction chromatography have been increasingly adopted and have been used as a more powerful substitute for SCX or SAX chromatography. HILIC is a variation of normal-mode chromatography in which a polar stationary phase and a non-polar or less polar mobile phase are used [35, 36]. In contrast to traditional normal-phase chromatography, an aqueous-organic mobile-phase mixture is used, typically a mixture of acetonitrile and water, where in contrast to reversed-phase LC the gradient elution is performed with decreasing acetonitrile concentration. Therefore, analytes are separated according to their hydrophilicity, with the more polar compounds eluting later. Phosphopeptides should therefore be enriched at the end of the gradient.

ERLIC is a variation of the HILIC method that uses an electrostatic repulsion effect to generate enhanced separation between phospho- and non-phosphopeptides [37]. For this purpose, an ion-exchange column is operated in a highly organic mobile phase (therefore resembling HILIC mode), so that several retention mechanisms—electrostatic attraction, hydrophilic interaction, and electrostatic repulsion are at play. Phosphopeptides are retained strongly under low pH, high acetonitrile conditions.

3.7 Other Methods

Different other recognition principles have been used to enrich phosphopeptides. For example, inorganic phosphates are known to form very insoluble salts with some counterions such as barium or calcium. This concept has been extended to phosphopeptide precipitation with Ca^{2+} [38], Ba^{2+} [39], or cations of the lanthanide group [40]. The insoluble precipitates are simply isolated by centrifugation and redissolved before LC-MS analysis.

Derivatization or tagging methods involve the chemical modification of the phosphate group to allow the introduction of affinity tags or the covalent binding to a stationary phase. The main concepts introduced for this purpose (reviewed in [41]) include the formation of phosphoamidates and the β-elimination of the phosphate group in phospho-Ser (pSer) and phospho-Thr (pThr) residues. The phosphoamidate chemistry is able to convert pSer, pThr, and pTyr peptides, but involves a multi-step reaction scheme to ensure specificity. Its application was restricted to a few studies from the Aebersold laboratory [42, 43].

β-Elimination is more commonly used and is based on the removal of the phosphate group from serine and threonine residues under strongly alkaline conditions. The resulting double bond in the dehydroamino acid residues can be reacted with various reagents in a Michael addition. However, reaction conditions

must be carefully optimized to avoid side reactions. A variety of different protocols have been proposed in the literature, most recently, the Angeletti group has studied experimental parameters in detail in several studies [44, 45], which were, however, restricted to less complex samples (see also protocol by Nika et al. in Chapter 3).

In addition to IMAC and MOAC materials, several other sorbents have been proposed for phosphopeptide enrichment, although many of them have only been evaluated in proof-of-principle experiments. Hydroxyl phosphates such as hydroxyapatite, $Ca_5(PO_4)_3(OH)$ [46], and $Fe_5(PO_4)_4(OH)_3$ [47] were shown to bind phosphopeptides preferentially, as were zirconium arsenate-modified nanoparticles [48] and a SnO_2-$ZnSn(OH)_6$ hybrid material [49]. Lanthanide [50] and yttrium [51] phosphates have also been proposed as affinity materials. Moreover, boron nitride [52, 53] has been proposed for the retention of very hydrophilic phosphopeptides, similar to HILIC materials.

3.8 Multidimensional Enrichment and Fractionation Methods

The different characteristics of the various methods introduced above and their (partial) complementarity have led to the development of multidimensional methods for phosphoproteomics. These integrated methods also address the point that a phosphopeptide mixture will remain highly complex even after a single enrichment step to an extent that often surpasses the sequencing capacity of current mass spectrometers. Thus, if sufficient amounts of starting material are available, various combinations of enrichment and fractionation strategies can be envisaged.

In the classical *two-dimensional* setup one of the less specific chromatographic fractionation procedures (SCX, SAX, HILIC, ERLIC) as the first stage may be combined with a metal affinity enrichment step (either IMAC or MOAC) as the second dimension. Although this is the most commonly used scenario, the inverse strategy may be more practical [54]. This is because phosphopeptides can be very unevenly distributed in different chromatographic fractions and using the same second-dimension enrichment protocol for all of them may sacrifice some of the performance.

Alternatively, *sequential* elution methods such as SIMAC (sequential elution from IMAC, [55], see protocol by Thingholm and Larsen in Chapter 10) may use two specific procedures in succession whereby the unbound fraction of the first stage is used as the starting material for the second stage. In the SIMAC method, a partial separation of singly and multiply phosphorylated peptides is achieved by the preferential elution of mono-phosphorylated peptides from an IMAC resin and the subsequent re-binding on TiO_2. Multiply phosphorylated peptides, on the other hand, are directly eluted from the IMAC resin. SIMAC has also been integrated into the multidimensional TiSH method that combines titania enrichment, SIMAC and HILIC fractionation ([56], See also protocol by Engholm-Keller and Larsen et al. in Chapter 11).

A third option is the *parallel* processing of peptide mixtures using complementary methods. This is particularly relevant for the comprehensive profiling of tyrosine phosphorylation, because the low abundance of pTyr-containing peptides usually leads to their underrepresentation even in phosphoenriched samples. Therefore, the majority of the sample is usually dedicated to enrichment with pTyr antibodies, while the rest is processed using metal affinity methods.

Finally, enrichment methods that integrate the profiling of different types of PTMs have also been described recently [57–59]. This opens up interesting perspectives into the regulation of biological signal processing using different types of modifications ("PTM cross talk," [60]).

4 Selected Applications

As outlined in the previous section, many different techniques for phosphoproteomics are employed at the sample processing stage. Picking the most suitable tool (or more often, a combination of tools) is therefore not an easy task. The rapid pace of instrument development and the diversity in methods (both experimental and computational) and sample types being analyzed do not allow for a straightforward comparison. Published reports may differ substantially in important parameters such as the number of fractions analyzed, total analysis time, number of MS/MS spectra, fragmentation method, database search engine, statistical thresholds, and many more. It is therefore unrealistic to highlight the "best" methods for the proteome-wide analysis of phosphorylation; instead I opted to highlight a few notable contributions to the field that have occurred over the last few years. This list (*see* Table 1) is by no means exhaustive and should not be seen as a comprehensive review of the field. Quantitative data (number of identifications) were taken directly from the papers and may be reported at different levels (phosphopeptides, phosphorylation sites) and at different levels of statistical confidence.

Over the course of roughly a decade, the field has moved from the identification of several hundred to tens of thousands of phosphopeptides in a single study. For example, the work by Ficarro et al. [61], who identified 386 phosphorylation sites on 216 peptides in yeast, was a substantial achievement at the time. For this the authors used single-stage IMAC enrichment after methyl esterification to reduce nonspecific binding and MS analysis on a low-resolution ion trap. Only a few years later, in 2006, Olsen et al. [33] already reported the identification of 6600 phosphorylation sites on more than 2200 proteins from EGF-treated HeLa cells. This study used SCX fractionation followed by TiO_2 enrichment and analysis of the fractions on a linear ion trap-Fourier transform

Table 1
Selected examples of phosphoproteomic studies from the recent literature (since 2010) that reported the identification of more than 5000 phosphopeptides or phosphorylation sites

Reference	Sample	Enrichment and fractionation	Mass spectrometry platform	Identifications (*)
Engholm-Keller et al. [56]	Rat cell line (INS-1)	TiO_2/SIMAC/HILIC	LTQ OT Velos (HCD)	Up to 7072 phosphopeptides
Fukuda et al. [27]	Human cell line (PC3)	HILIC/TiO_2	Q Exactive (HCD)	Up to 6841 phosphopeptides, 2387 proteins
Hennrich et al. [63]	Human cell line (HEK293)	SCX/WAX	LTQ OT Velos (HCD)	Up to 11010 peptides
Hennrich et al. [64]	Human cell line (HeLa)	SCX (at pH 1 and pH 3)	LTQ OT Velos (HCD, ETD)	9673 phosphopeptides
Huttlin et al. [65]	Mouse (9 tissues)	SCX/Fe^{3+}-IMAC	LTQ OT Velos (CID)	35965 phosphorylation sites, 6296 proteins
Iesmantavicius et al. [59]	Yeast	TiO_2/SCX and anti-Gly-Gly-antibody/SCX in parallel	Q Exactive (HCD)	8961 phosphorylation sites, 2498 ubiquitination sites
Lundby et al. [66]	Rat (14 organs and tissues)	TiO_2	LTQ OT Velos (HCD)	Total 31480 phosphorylation sites, 7280 proteins
Meijer et al. [67]	Human cell line (HUVEC)	SCX/Ti^{4+}-IMAC	LTQ OT Velos (HCD, ETD)	19859 phosphopeptides, 17278 phosphorylation sites, 4594 proteins
Mertins et al. [57]	Human cell line (Jurkat)	Fe^{3+}-IMAC/anti-GG-antibody/anti-KAc antibody	Q Exactive (HCD)	Up to 20800 phosphorylation sites, 15408 ubiquitination sites, 3190 KAc sites
Monetti et al. [68]	Mouse liver	SCX/TiO_2	LTQ OT Velos (HCD)	Up to 20491 phosphorylation sites
Nagaraj et al. [69]	Human cell line (HeLa)	SCX/TiO_2	LTQ OT Velos (HCD or CID)	Up to 9668 high-confidence phosphorylation sites
Olsen et al. [70]	Human cell line (HeLa)	SCX/TiO_2 and others	LTQ Orbitrap (MSA)	20443 phosphorylation sites, 6027 proteins
Phanstiel et al. [71]	Human stem cells	SCX/Fe^{3+}-IMAC	LTQ OT Velos	19122 phosphorylation sites
Rigbolt et al. 2011 [72]	Human stem cells	SCX/TiO_2 and others	LTQ Orbitrap and others	23522 phosphorylation sites, 6521 proteins

(continued)

Table 1 (continued)

Reference	Sample	Enrichment and fractionation	Mass spectrometry platform	Identifications (*)
Yi et al. [73]	Human stem cells	SCX/Fe^{3+}-IMAC	LTQ OT Velos (CID)	11131 phosphorylation sites, 2567 proteins
Yue et al. [74]	Human cell line (MCF-10A)	Multi-step IMAC/high-pH reversed-phase fractionation	Q Exactive (HCD)	8969 phosphopeptides
Zarei et al. 2012 [75]	Human cell line (HeLa)	ERLIC/SCX + TiO_2 in different set-ups	LTQ Orbitrap XL (CID)	Up to 13585 phosphopeptides
Zhou et al. 2013 [76]	Human cell line (HeLa, K562)	SCX/Ti^{4+}-IMAC/HILIC in different set-ups	LTQ OT Velos (HCD, ETD)	Up to 22148 phosphopeptides, 18055 phosphorylation sites, 4708 proteins

Studies are sorted alphabetically after the first author. Abbreviations: *CID* collision-induced dissociation, *ERLIC* electrostatic repulsion hydrophilic interaction chromatography, *ETD* electron transfer dissociation, *HCD* higher collisional energy dissociation, *HILIC* hydrophilic interaction chromatography, *IMAC* immobilized metal ion affinity chromatography, *KAc* acetyl-lysine, *MSA* multistage activation, *OT* Orbitrap, *SCX* strong cation exchange, *SIMAC* sequential elution from IMAC, *WAX* weak anion exchange. (*) "up to" means that numbers depend on actual sample, degree of fractionation, etc.

ion cyclotron resonance mass spectrometer. Therefore, a combination of increased analysis time, a faster scanning and more sensitive instrument, and of course the application to a different organism with a significantly higher number of phosphorylation events resulted in more comprehensive phosphoproteome coverage.

More recently, different comprehensive studies have allowed the identification of more than 10,000 phosphopeptides or phosphorylation sites from various sample types. Some of the key method parameters of such studies are compared and summarized in Table 1, although for more specific information the readers are referred to the original articles. A few common concepts become obvious upon comparison of the different approaches. First of all, all studies used Orbitrap mass spectrometers of various generations [62], highlighting the Orbitrap as the instrument platform of choice for phosphoproteomics, similar to general proteomics methods. Secondly, SCX is widely used in combination with IMAC or TiO_2-MOAC, despite its known limitations, especially the limited chromatographic resolution of the technique. However, alternative workflows such as HILIC and ERLIC are also capable of achieving similar numbers of identifications, and even single-stage enrichment methods are highly promising.

While the numbers reported from the sophisticated studies summarized here are certainly impressive, it has to be mentioned that such results are achievable if large sample amounts (frequently in the mg range) and, frequently, considerable instrument time (>1 day) are available for each sample. In contrast, emerging clinical applications of phosphoproteomic methods will require substantially higher sample throughput, making extensive fractionation impractical. In addition, the limited sample amounts require the development of optimized sample preparation procedures, especially those that minimize sample handling. For example, Zou and co-workers recently introduced an integrated method combining sequential protein digestion, differential stable isotope labeling, and Ti^{4+}-IMAC enrichment of phosphopeptides in the same tube for the processing of small amounts of tissue lysates [77].

5 Conclusion and Outlook

The biological relevance of protein phosphorylation and the successful application of mass spectrometry-based proteomics workflows have resulted in a wide variety of methods and protocols that can be used to study the phosphoproteome. Although the majority of methods are directed to the analysis of serine and threonine (and to some extent, tyrosine) phosphorylation, less abundant targets are gaining increasing interest. Considering the ever improving performance of mass spectrometers, it is reasonable to expect that in the coming years the phosphoproteome can be probed to an even greater depth. Robust and reproducible enrichment methods will contribute to an increased application of phosphoproteomics also for biomedical and clinical applications, where sample amounts are frequently very limited.

References

1. Kim M-S, Pinto SM, Getnet D et al (2014) A draft map of the human proteome. Nature 509:575–581
2. Wilhelm M, Schlegl J, Hahne H et al (2014) Mass-spectrometry-based draft of the human proteome. Nature 509:582–587
3. Beltran L, Cutillas PR (2012) Advances in phosphopeptide enrichment techniques for phosphoproteomics. Amino Acids 43:1009–1024
4. Fila J, Honys D (2012) Enrichment techniques employed in phosphoproteomics. Amino Acids 43:1025–1047
5. Engholm-Keller K, Larsen MR (2013) Technologies and challenges in large-scale phosphoproteomics. Proteomics 13:910–931
6. Loroch S, Dickhut C, Zahedi RP, Sickmann A (2013) Phosphoproteomics—More than meets the eye. Electrophoresis 34:1483–1492
7. Guo M, Huang BX (2013) Integration of phosphoproteomic, chemical, and biological strategies for the functional analysis of targeted protein phosphorylation. Proteomics 13: 424–437
8. Roux PP, Thibault P (2013) The coming of age of phosphoproteomics—from large data sets to inference of protein functions. Mol Cell Proteomics 12:3453–3464
9. Jünger MA, Aebersold R (2014) Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. Wiley Interdiscip Rev Dev Biol 3:83–112

10. Wells JM, McLuckey SA (2005) Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol 402:148–185
11. Boersema PJ, Mohammed S, Heck AJR (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. J Mass Spectrom 44: 861–878
12. Chalkley RJ, Clauser KR (2012) Modification site localization scoring: strategies and performance. Mol Cell Proteomics 11:3–14
13. Rush J, Moritz A, Lee KA et al (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 23: 94–101
14. Boersema PJ, Foong LY, Ding VMY et al (2010) In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol Cell Proteomics 9:84–99
15. Kee J-M, Oslund RC, Perlman DH et al (2013) A pan-specific antibody for direct detection of protein histidine phosphorylation. Nat Chem Biol 9:416–421
16. Porath J, Carlsson J, Olsson I et al (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258:598–599
17. Andersson L, Porath J (1986) Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Anal Biochem 154:250–254
18. Zhou H, Ye M, Dong J et al (2008) Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. J Proteome Res 7:3957–3967
19. Zhou H, Ye M, Dong J et al (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. Nat Protoc 8:461–480
20. Iliuk AB, Martin VA, Alicie BM et al (2010) In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. Mol Cell Proteomics 9:2162–2172
21. Jayasundera KB, Iliuk AB, Nguyen A et al (2014) Global phosphoproteomics of activated b cells using complementary metal ion functionalized soluble nanopolymers. Anal Chem 86:6363–6371
22. Tsai C-F, Hsu C-C, Hung J-N et al (2014) Sequential phosphoproteomic enrichment through complementary metal-directed immobilized metal ion affinity chromatography. Anal Chem 86:685–693
23. Pinkse MWH, Uitto PM, Hilhorst MJ et al (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76:3935–3943
24. Larsen MR, Thingholm TE, Jensen ON et al (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4:873–886
25. Leitner A (2010) Phosphopeptide enrichment using metal oxide affinity chromatography. Trends Anal Chem 29:177–185
26. Sugiyama N, Masuda T, Shinoda K et al (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
27. Fukuda I, Hirabayashi-Ishioka Y, Sakikawa I et al (2013) Optimization of enrichment conditions on TiO2 chromatography using glycerol as an additive reagent for effective phosphoproteomic analysis. J Proteome Res 12:5587–5597
28. Kyono Y, Sugiyama N, Imami K et al (2008) Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography. J Proteome Res 7:4585–4593
29. Imami K, Sugiyama N, Kyono Y et al (2008) Automated phosphoproteome analysis for cultured cancer cells by two-dimensional NanoLC-MS using a calcined titania/C18 Biphasic column. Anal Sci 24:161–166
30. Li Q-R, Ning Z-B, Tang J-S et al (2009) Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. J Proteome Res 8:5375–5381
31. Schmidt A, Trentini DB, Spiess S et al (2014) Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. Mol Cell Proteomics 13:537–550
32. Beausoleil SA, Jedrychowski M, Schwartz D et al (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101:12130–12135
33. Olsen JV, Blagoev B, Gnad F et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648
34. Han G, Ye M, Zhou H et al (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics 8:1346–1361

35. Boersema PJ, Mohammed S, Heck AJR (2008) Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem 391:151–159
36. McNulty DE, Annan RS (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Mol Cell Proteomics 7:971–980
37. Alpert AJ (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem 80:62–76
38. Zhang X, Ye J, Jensen ON et al (2007) Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. Mol Cell Proteomics 6:2032–2042
39. Ruse CI, McClatchy DB, Lu B et al (2008) Motif-specific sampling of phosphoproteomes. J Proteome Res 7:2140–2150
40. Güzel Y, Rainer M, Mirza MR et al (2013) Highly selective recovery of phosphopeptides using trypsin-assisted digestion of precipitated lanthanide-phosphoprotein complexes. Analyst 138:2897–2905
41. Leitner A, Lindner W (2009) Chemical tagging strategies for mass spectrometry-based phospho-proteomics. Methods Mol Biol 527:229–243
42. Zhou H, Watts JD, Aebersold R (2001) A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol 19:375–378
43. Tao WA, Wollscheid B, O'Brien R et al (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. Nat Methods 2:591–598
44. Nika H, Lee J, Willis IM et al (2012) Phosphopeptide characterization by mass spectrometry using reversed-phase supports for solid-phase β-elimination/michael addition. J Biomol Tech 23:51–68
45. Nika H, Nieves E, Hawke DH et al (2013) Optimization of the β-elimination/michael addition chemistry on reversed-phase supports for mass spectrometry analysis of O-linked protein modifications. J Biomol Tech 24:132–153
46. Fonslow BR, Niessen SM, Singh M et al (2012) Single-step inline hydroxyapatite enrichment facilitates identification and quantitation of phosphopeptides from mass-limited proteomes with MudPIT. J Proteome Res 11:2697–2709
47. Chen Q, Wei C, Zhang Y et al (2014) Single-crystalline hyperbranched nanostructure of iron hydroxyl phosphate $Fe_5(PO_4)_4(OH)_3$ x 2 H_2O for highly selective capture of phosphopeptides. Sci Rep 4:3753
48. Li X-S, Xu L-D, Zhu G-T et al (2012) Zirconium arsenate-modified magnetic nanoparticles: preparation, characterization and application to the enrichment of phosphopeptides. Analyst 137:959–967
49. Li L-P, Zheng T, Xu L-N et al (2013) SnO_2-$ZnSn(OH)_6$: a novel binary affinity probe for global phosphopeptide detection. Chem Commun 49:1762–1764
50. Cheng G, Li S-M, Wang Y et al (2013) $REPO_4$ (RE = La, Nd, Eu) affinity nanorods modified on a MALDI plate for rapid capture of target peptides from complex biosamples. Chem Commun 49:8492–8494
51. Cheng G, Liu Y-L, Wang Z-G et al (2013) Yolk-shell magnetic microspheres with mesoporous yttrium phosphate shells for selective capture and identification of phosphopeptides. J Mater Chem B 1:3661–3669
52. Fischnaller M, Köck R, Bakry R et al (2014) Enrichment and desalting of tryptic protein digests and the protein depletion using boron nitride. Anal Chim Acta 823:40–50
53. Furuhashi T, Nukarinen E, Ota S et al (2014) Boron nitride as desalting material in combination with phosphopeptide enrichment in shotgun proteomics. Anal Biochem 452:16–18
54. Li Q-R, Ning Z-B, Yang X-L (2012) Complementary workflow for global phosphoproteome analysis. Electrophoresis 33:3291–3298
55. Thingholm TE, Jensen ON, Robinson PJ et al (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. Mol Cell Proteomics 7:661–671
56. Engholm-Keller K, Birck P, Storling J et al (2012) TiSH—a robust and sensitive global phosphoproteomics strategy employing a combination of TiO_2, SIMAC, and HILIC. J Proteomics 75:5749–5761
57. Mertins P, Qiao JW, Patel J et al (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods 10:634–637
58. Swaney DL, Beltrao P, Starita L et al (2013) Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods 10:676–682
59. Iesmantavicius V, Weinert BT, Choudhary C (2014) Convergence of ubiquitylation and phosphorylation signaling in rapamycin-treated yeast cells. Mol Cell Proteomics 13:1979–1992
60. Beltrao P, Bork P, Krogan NJ et al (2013) Evolution and functional cross-talk of protein post-translational modifications. Mol Syst Biol 9:714

61. Ficarro SB, McCleland ML, Stukenberg PT et al (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol 20:301–305
62. Zubarev RA, Makarov A (2013) Orbitrap mass spectrometry. Anal Chem 85:5288–5296
63. Hennrich ML, Groenewold V, Kops GJ, Heck AJ et al (2011) Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach. Anal Chem 83:7137–7143
64. Hennrich ML, van den Toorn HWP, Groenewold V et al (2012) Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. Anal Chem 84:1804–1808
65. Huttlin EL, Jedrychowski MP, Elias JE et al (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143: 1174–1189
66. Lundby A, Secher A, Lage K et al (2012) Quantitative maps of protein phosphorlyation sites across 14 different rat organs and tissues. Nat Commun 3:876
67. Meijer LAT, Zhou H, Chan OYA et al (2013) Quantitative global phosphoproteomics of human umbilical vein endothelial cells after activation of the Rap signaling pathway. Mol Bio Syst 9:732–749
68. Monetti M, Nagaraj N, Sharma K et al (2011) Large-scale phosphosite quantification in tissues by a spike-in SILAC method. Nat Methods 8:655–658
69. Nagaraj N, D'Souza RCJ, Cox J et al (2010) Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. J Proteome Res 9:6786–6794
70. Olsen JV, Vermeulen M, Santamaria A et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3:ra3
71. Phanstiel DH, Brumbaugh J, Wenger CD et al (2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat Methods 8:821–827
72. Rigbolt KTG, Prokhorova TA, Akimov V et al (2011) System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal 4:rs3
73. Yi T, Zhai B, Yu Y et al (2014) Quantitative phosphoproteomic analysis reveals system-wide signaling pathways downstream of SDF-1/CXCR4 in breast cancer stem cells. Proc Natl Acad Sci U S A 111:E2182–E2190
74. Yue X-S, Hummon AB (2013) Combination of Multistep IMAC Enrichment with High-pH Reverse Phase Separation for In-Depth Phosphoproteomic Profiling. J Proteome Res 12:4176–4186
75. Zarei M, Sprenger A, Gretzmeier C et al. (2012) Combinatorial Use of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) and Strong Cation Exchange (SCX) Chromatography for In-Depth Phosphoproteome Analysis. J Proteome Res 12: 4269–4276
76. Zhou H, Di Palma S, Preisinger C et al. (2013) Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. J Proteome Res 13:260–271
77. Huang J, Qin H, Dong J et al (2014) In Situ Sample Processing Approach (iSPA) for comprehensive quantitative phosphoproteome analysis. J Proteome Res 13:3896–3904

Chapter 8

Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography

Tine E. Thingholm and Martin R. Larsen

Abstract

Immobilized metal affinity chromatography (IMAC) has been the method of choice for phosphopeptide enrichment prior to mass spectrometric analysis for many years and it is still used extensively in many laboratories. Using the affinity of negatively charged phosphate groups towards positively charged metal ions such as Fe^{3+}, Ga^{3+}, Al^{3+}, Zr^{4+}, and Ti^{4+} has made it possible to enrich phosphorylated peptides from peptide samples. However, the selectivity of most of the metal ions is limited, when working with highly complex samples, e.g., whole-cell extracts, resulting in contamination from nonspecific binding of non-phosphorylated peptides. This problem is mainly caused by highly acidic peptides that also share high binding affinity towards these metal ions. By lowering the pH of the loading buffer nonspecific binding can be reduced significantly, however with the risk of reducing specific binding capacity. After binding, the enriched phosphopeptides are released from the metal ions using alkaline buffers of pH 10–11, EDTA, or phosphate-containing buffers.

Here we describe a protocol for IMAC using Fe^{3+} for phosphopeptide enrichment. The principles are illustrated on a semi-complex peptide mixture.

Key words Protein phosphorylation, Phosphopeptide enrichment, Immobilized metal affinity chromatography, Mass spectrometry

1 Introduction

A widely used affinity enrichment technique for phosphorylated peptides is the use of metal ions for the binding of negatively charged phosphopeptides (e.g., immobilized metal affinity chromatography, IMAC). The affinity of proteins and peptides for metal ions due to certain amino acids (mainly histidine and cysteine residues) have been used for the purification of entire proteins [1, 2], but the binding of phosphoproteins and phosphoamino acids to metal ions, introduced by Andersson and Porath [3], gave the technique an extra dimension. With the work performed by

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_8,

Neville, D. C. et al. [4] on phosphopeptides from digested proteins the technique was further developed for the use on samples of low amount of starting material. The IMAC technique has since then been used extensively for enrichment of phosphorylated peptides prior to mass spectrometric analysis [5–10].

The technique improves identification of phosphopeptides from complex biological mixtures [8–10]. However, non-phosphorylated peptides containing multiple acidic amino acid residues co-purify with the phosphopeptides in IMAC. This causes unwanted high levels of non-phosphorylated peptides which ionize much better than phosphorylated peptides. Therefore, in order to decrease the number of non-phosphorylated peptides in large-scale phosphoproteomic studies O-methylesterification of the acidic amino acid residues has been shown to decrease the co-purification of acidic peptides [9]. This step, however, may introduce undesirable side reactions and loss of peptides due to adsorptive losses from extensive lyophilization [11]. Another possibility is to lower the complexity of the samples by performing pre-fractionation of the peptide mixture prior to IMAC using for example isoelectric focusing (e.g., [12]), ion-exchange chromatography (e.g., [8]), or hydrophilic interaction liquid chromatography (HILIC) [13] prior to MS analysis.

In 1996 Saha and co-workers published results showing that the p*K*a value of phosphoric acid decreased to 1.1 upon methylation [14]. As the addition of an organic group decreases the p*K*a value of phosphoric acid, it would be a reasonable assumption that the p*K*a values of phosphopeptides would be significantly lower than that of phosphoric acid due to the organic environment provided by the surrounding amino acids. By decreasing the pH of the loading conditions to below 1.9, more acidic peptides in a sample will become neutralized while phosphopeptides will retain their negative charges and their binding affinity towards the metal ions. By loading the sample in 0.1 % TFA and 50 % acetonitrile, the level of nonspecific binding to the IMAC material is significantly reduced [15]. Recently, an IMAC-Ti^{4+} resin has been designed [16, 17], which apparently has similar high tolerance towards acidic solutions and low nonspecific binding [18] compared to the TiO_2 enrichment described in Chapter 9.

Phosphopeptides are subsequently eluted from the IMAC material using alkaline buffers such as ammonia water (pH 10–11). Alternatively, phosphopeptides can be eluted with EDTA [19], highly acidic solutions [20] or solutions including phosphate or phosphoric acid [15, 21]. The eluted phosphopeptides should be desalted and concentrated by reversed phase micro-columns prior to downstream analysis.

2 Materials

2.1 Model Proteins

1. Transferrin (human) was a gift from ACE Biosciences A/S. Serum albumin (bovine), beta-lactoglobulin (bovine), carbonic anhydrase (bovine), beta-casein (bovine), alpha-casein (bovine), ovalbumin (chicken), ribonuclease B (bovine pancreas), alcohol dehydrogenase (Baker yeast), myoglobin (whale skeletal muscle), lysozyme (chicken), and alpha-amylase (bacillus species) were from Sigma (St. Louis. MO, US).

2.2 Reduction, Alkylation, and Digestion of Model Proteins

1. Ammonium bicarbonate.
2. Dithiothreitol (DTT).
3. Iodoacetamide.
4. Modified trypsin.

2.3 Immobilized Metal Ion Affinity Chromatography

1. Iron-coated PHOS-select™ metal chelate beads (Sigma®), stored at –20° (or other IMAC material with similar capacity).
2. IMAC loading buffer: 0.1 % trifluoroacetic acid (TFA), Protein Sequencer Grade, 50 % acetonitrile, HPLC grade.
3. GELoader tips (Eppendorf (20 μL) or Bio-Rad (200 μL)).
4. Low-binding microcentrifuge tubes 1.7 mL.
5. 1–5 mL disposable syringes fitted to GELoader tip or p200 tips by using a pipette tip cut in both ends.
6. IMAC Elution Buffer: 1 % ammonia water (40 μL ammonia solution (25 %), 980 μL UHQ water (pH ~11)), make fresh as required.
7. Formic acid.
8. Milli-Q water (UHQ water) (*see* **Notes 1** and **2**).

2.4 Reversed-Phase (RP) Chromatography

1. POROS Oligo R3 reversed-phase material (PerSeptive Biosystems, Framingham, MA, US).
2. GELoader tip (Eppendorf, Hamburg, Germany) or p200 pipette tips.
3. Syringe for HPLC loading (P/N 038250, N25/500-LC PKT 5, SGE, Ringwood, Victoria, Australia).
4. RP loading buffer: 0.1 % TFA.
5. RP elution buffer (for LC-ESI MS/MS analysis): 70 % acetonitrile, 0.1 % TFA.
6. 2,5-Dihydroxybenzoic acid (DHB) elution buffer (for MALDI MS analysis): 20 mg/mL DHB in 50 % acetonitrile, 0.1 % TFA, 1 % ortho-phosphoric acid.

2.5 Other Materials

1. Tabletop centrifuge.
2. pH meter.

3. Thermomixer.
4. Shaker.
5. Vacuum centrifuge.

2.6 Analysis by Mass Spectrometry

1. Mass spectrometer capable of performing MS/MS—preferentially a high-resolution/high-mass-accuracy instrument (Q-TOFs or Orbitrap-based mass spectrometer) interfaced to a nanoHPLC with a 50–100 μm i.d. RP capillary column setup for highly sensitive online peptide separation.
2. Software for processing of raw mass spectrometry data files and generation of peak lists for searching against a protein database (e.g., Uniprot) Analysis software such as Mascot/Mascot Distiller (Matrix Science, London, UK) (data from most vendors and instruments), Proteome Discoverer (ThermoScientific, Bremen, Germany) (data from Thermo instruments), MaxQuant [15] (high-resolution data from Thermo Orbitrap instruments and certain Bruker and ABSciex Q-TOFs) and the TransProteomicPipeline [16] (vendor independent).

3 Methods

The principle of the protocol shown in this chapter is illustrated using a semi-complex peptide mixture originating from tryptic digestions of 12 standard proteins (model proteins) (*see* **Notes 3** and **4**). The IMAC purification method is a simple and "easy-to-do" method. The buffers used for batch incubation with IMAC material should contain 0.1 % TFA and 50 % acetonitrile to optimize adsorption of phosphopeptides to the IMAC beads and reduce nonspecific binding [15]. The commonly used ammonium bicarbonate buffer should be avoided (*see* **Note 5**).

3.1 Model Proteins and Peptide Mixture

1. Dissolve each protein in 50 mM triethylammonium bicarbonate (TEAB), pH 7.8, and 10 mM DTT and incubate at 37 °C for 1 h. After reduction, add 20 mM iodoacetamide and incubate the samples at room temperature for 1 h in the dark.
2. Digest each protein using trypsin (1–2 % w/w) at 37 °C for 12 h.
3. Quench the reaction by adding formic acid to a final concentration of 2 %.

3.2 Batch Incubation with IMAC Beads

1. When working with low amounts of sample use less IMAC beads to reduce the level of nonspecific binding from non-phosphorylated peptides. For 1 pmol tryptic digest use 7 μL IMAC beads. For more complex samples, where more material is available, more IMAC beads should be used. In the

following we will illustrate the method for 1 pmol tryptic digest of the protein mixture.

2. Transfer 7 µL IMAC beads to a fresh 1.7 mL low binding microcentrifuge tube.
3. Wash the IMAC beads twice using 50 µL IMAC Loading Buffer (*see* **Note 6**).
4. Resuspend the beads in 40 µL IMAC Loading Buffer and add the sample (*see* **Note 7**).
5. Incubate the sample with IMAC beads in a Thermomixer for 30 min at room temperature.

3.3 Packing the IMAC Micro-Column

1. Squeeze the tip of a GELoader tip to prevent the IMAC beads from leaking.
2. Pack the beads in the constricted end of the GELoader tip by application of air pressure forming an IMAC micro-column [22] (*see* Fig. 1). For larger amount of IMAC beads and more complex mixtures a p200 GELoader tip is recommended.
3. Wash the IMAC column using 40 µL IMAC Loading Buffer.

3.4 Elution of Phosphorylated Peptides from the IMAC Micro-Column

1. Elute the phosphorylated peptides bound to the IMAC micro-column using 30 µL of IMAC elution buffer (*see* Fig. 2). It is important that this step is performed slowly (~1 drop/s). (N.B. For MALDI MS analysis the peptides can be eluted off the IMAC micro-column directly onto the MALDI target using 1 µL DHB solution. After crystallization the sample is ready for MALDI MS analysis) (*see* **Note 8**).
2. For LC-ESI MS/MS analysis the IMAC eluent should be acidified using 100 % formic acid (pH should be ~2–3) and desalted/concentrated using reversed-phase micro-columns.

3.5 Poros Oligo R3 Reversed-Phase (RP) Micro-Column Desalting/ Concentration of the Sample

Use RP GELoader tip micro-columns of ~6–10 mm or p200 pipette tips micro-columns (1–2 cm) depending on the amount of material to be purified. Here, it is illustrated for a GELoader tip micro-column.

1. Suspend Poros Oligo R3 reversed-phase (RP) material in 200 µL 100 % acetonitrile.
2. Squeeze the tip of a GELoader tip to prevent the RP beads from leaking (*see* Fig. 1).
3. Pack the beads in the constricted end of the GELoader tip by application of air pressure forming a micro-column [22].
4. Load the acidified phosphopeptide sample slowly onto the RP micro-column (~1 drop/s).
5. Wash the RP micro-column using 30 µL RP loading buffer.

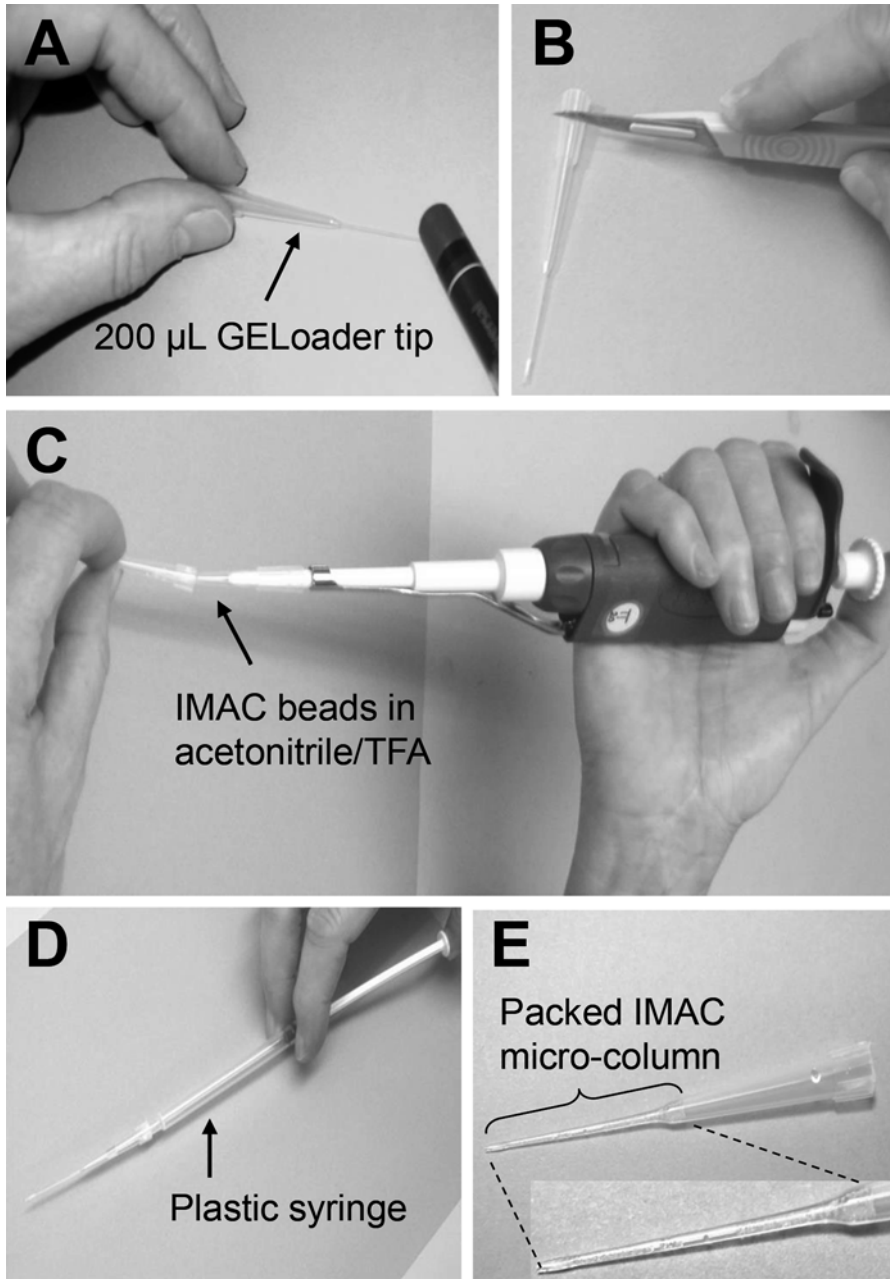


Fig. 1 The principle of making IMAC micro-columns for a complex sample. The constricted end of a 200 µL GELoader tip is squeezed to prevent the IMAC beads from leaking. This is done by pressing the tip of a pen onto the end of the GELoader tip (**a**). Cut the top of the GELoader tip to make a plastic syringe fit into the opening (**b**). Load the IMAC beads onto the GELoader tip (**c**). Pack the IMAC beads to form an IMAC micro-column by applying air pressure using a 1 mL plastic syringe (**d**). The packed IMAC micro-column (**e**). The principle for a simple sample is the same, however, using a 20 µL Eppendorf GELoader tip instead of a 200 µL GELoader tip. This figure is taken from [25]

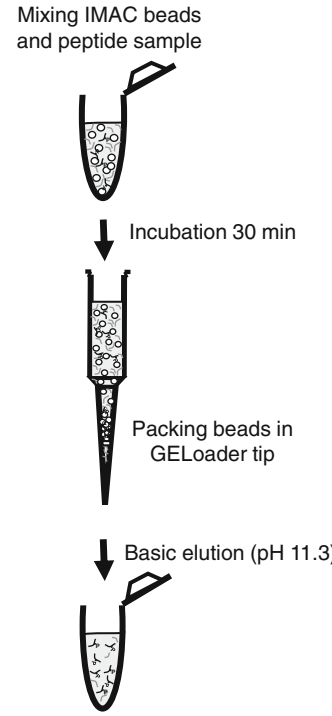


Fig. 2 Illustration of the IMAC strategy

6. Elute the phosphopeptides from the RP micro-column using 20 µL RP Elution Buffer, followed by lyophilization of the phosphopeptides. (N.B. For MALDI MS analysis the peptides can be eluted off the GELoader tip RP micro-column directly onto the MALDI target using 1 µL DHB solution. After crystallization the sample is ready for MALDI MS analysis.)
7. Redissolve the dried phosphopeptides in 0.5 µL 100 % formic acid and dilute immediately to 10 µL with UHQ water. The sample is then ready for LC-ESI-MS/MS analysis.

3.6 µHPLC Tandem Mass Spectrometry (LC-MS/MS) Analysis

For LC-MS/MS analysis of purified phosphopeptides a standard strategy as described below can be used. A typical nanoLC setup would include a 0.075 mm × 200 mm analytical column packed with 3 µm RP resin interfaced with a high-resolution/mass accuracy mass spectrometer as described in our original paper [23]. The number of phosphopeptides identified in the analysis can be increased by maximizing the resolution of the nanoLC separation via longer columns (e.g., 50 cm) and smaller chromatographic particle sizes (e.g., 1.9 µm). Alternatively, a two column system can be utilized using a 0.1 mm × 20 mm pre-column packed with RP resin (3–5 µm) combined with an analytical column as described above. A two-column system is described below.

1. The phosphopeptides are redissolved in 0.1 % TFA and loaded onto a pre-column as described above using a μHPLC system (e.g., Dionex or EASY-LC) at a loading speed of 5 μL/min.
2. The phosphopeptides are eluted directly onto the analytical column (e.g., 0.075 mm × 200 mm) using a gradient (60–120 min) from 0 to 35 % B-Buffer (e.g., A-Buffer: 0.1 % formic acid; B-Buffer: 90 % acetonitrile, 0.1 % TFA) at an elution speed of 2–300 nL/min.
3. The phosphopeptides are eluted directly into a tandem mass spectrometer and analyzed by Data Dependent Analysis.

LC-ESI-MS/MS analysis of multi-phosphorylated peptides is improved by redissolving the phosphopeptides by sonication in an EDTA containing buffer prior to LC-ESI-MS/MS analysis [24].

An example of the results obtained when using this procedure for a relative low complexity sample is shown in Fig. 3. The figure illustrates the analysis of 1 pmol peptide mixture by MALDI MS as a dried droplet without IMAC enrichment (a), the MALDI MS

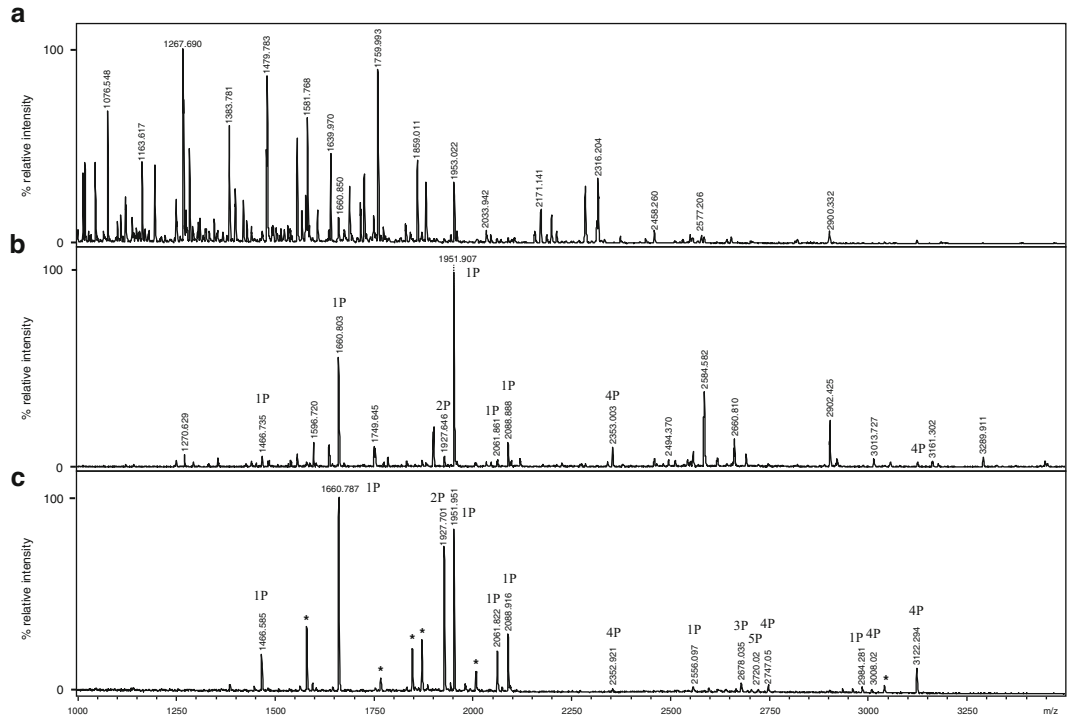


Fig. 3 Results obtained from 1 pmol peptide mixture using the IMAC strategy. (**a**) MALDI MS peptide mass map of the direct analysis of the tryptic peptides. (**b**) MALDI MS peptide mass map of phosphopeptides purified by IMAC from the standard peptides mixture using 0.1 M acetic acid and 30 % acetonitrile (pH 2.9) as loading buffer. (**c**) MALDI MS peptide mass map of phosphopeptides purified by IMAC from the standard peptides mixture using 0.1 % TFA and 50 % acetonitrile (pH 1.8) as loading buffer. The number of phosphate groups on the individual phosphopeptides is indicated by "#P." *Asterisk* indicates the metastable loss of phosphoric acid

peptide mass map obtained from 1 pmol peptide mixture after IMAC phosphopeptide enrichment using 0.1 M acetic acid (b) or 0.1 % TFA (c) as loading buffer (*see* **Notes 9** and **10**). For analysis of this sample using LC-ESI-MS/MS the above strategy can be used (*see* **Note 11**).

4 Notes

1. It is important to obtain the highest purity of all chemicals used.
2. All solutions should be prepared in UHQ water.
3. Always start by testing the method using a model peptide mixture. It is important to freshly prepare the peptide mixture as peptides bind to the surface of the plastic tubes in which they are stored. In addition, avoid transferring the peptide sample between different tubes to minimize adsorptive losses of the sample.
4. The peptide mixture used for the experiment illustrated in this chapter contained peptides originating from tryptic digestions of 1 pmol of each of the 12 proteins.
5. If the sample has been pre-suspended in another buffer check whether the buffer may interfere with IMAC binding [19]. For example, the use of the widely used ammonium bicarbonate buffer reduces the binding of phosphopeptides to IMAC material up to 75 %, presumable due to shielding of the phosphate groups which prevent interaction with the IMAC resin (Larsen MR unpublished results). Alternatively, triethylammonium bicarbonate (TEAB) should be used. If the sample contains other reagents, which interfere with IMAC binding, dilute the samples sufficiently to reduce the concentration of these reagents or the samples should be desalted. If there is EDTA present in the sample the peptides have to be purified using reversed phase chromatography prior to IMAC. Always test the pH value of the sample before IMAC batch incubation. The pH value should be approximately 1.7.
6. The PhosSelect IMAC beads are very fragile, so high-speed mixing should be avoided in all steps.
7. When working with larger amounts of IMAC beads remember to also increase the incubation, washing and elution volumes.
8. When working with larger amounts of material it is better to elute the phosphopeptides using IMAC elution buffer followed by acidification and reversed phase chromatography, since 1 μL DHB buffer will not be sufficient to elute all phosphopeptides and LC-MS/MS cannot be performed in the presence of DHB. In addition, the sample may be too

concentrated for MS analysis and only a fraction of the eluent needs to be analyzed.

9. The results obtained using this protocol will differ according to the mass spectrometer used for the analysis of the phosphopeptides, not only between MALDI MS and ESI MS but also within different MALDI MS instruments, depending on laser optics, laser frequency, instrumental configuration, sensitivity, etc.
10. The exact binding affinity of the IMAC beads is not known; however, the amount of nonspecific binding from non-phosphorylated peptides is very dependent on the ratio between amount of sample and IMAC beads. It may be necessary to optimize the ratio for different samples.
11. Analysis of multiply phosphorylated peptides using LC-ESI-MS/MS can be complicated and peptides with >3 phosphate groups are rarely observed using this method. Thus MALDI MS analysis of known phosphopeptides is a better choice for setting up and testing methods for phosphopeptide enrichment.

Acknowledgements

This work was supported by the Danish Natural Science and Medical Research Councils (grant no. 10-082195 (T.E.T.)) and the Lundbeck Foundation (M.R.L.—Junior Group Leader Fellowship).

References

1. Chaga G, Hopp J, Nelson P (1999) Immobilized metal ion affinity chromatography on Co2+-carboxymethylaspartate–agarose Superflow, as demonstrated by one-step purification of lactate dehydrogenase from chicken breast muscle. Biotechnol Appl Biochem 29:19–24
2. Hochuli E, Dobeli H, Schacher A (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. J Chromatogr 411:177–184
3. Andersson L, Porath J (1986) Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Anal Biochem 154:250–254
4. Neville DC, Rozanas CR, Price EM, Gruis DB, Verkman AS, Townsend RR (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. Protein Sci 6:2436–2445
5. Figeys D, Gygi SP, McKinnon G, Aebersold R (1998) An integrated microfluidics-tandem mass spectrometry system for automated protein analysis. Anal Chem 70:3728–3734
6. Li S, Dass C (1999) Iron(III)-immobilized metal ion affinity chromatography and mass spectrometry for the purification and characterization of synthetic phosphopeptides. Anal Biochem 270:9–14
7. Posewitz MC, Tempst P (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. Anal Chem 71:2883–2892
8. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101:12130–12135
9. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to

Saccharomyces cerevisiae. Nat Biotechnol 20:301–305

10. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics 4:310–327
11. Stewart II, Thomson T, Figeys D (2001) O-18 Labeling: a tool for proteomics. Rapid Commun Mass Spectrom 15:2456–2465
12. Krijgsveld J, Gauci S, Dormeyer W, Heck A (2006) In-gel isoelectric focusing of peptides as a tool for improved protein identification. J Proteome Res 5:1721–1730
13. McNulty DE, Annan RS (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Mol Cell Proteomics 7:971–980
14. Saha A, Saha N, Ji LN, Zhao J, Gregan F, Sajadi SAA, Song B, Sigel H (1996) Stability of metal ion complexes formed with methyl phosphate and hydrogen phosphate. J Biol Inorg Chem 1:231–238
15. Kokubu M, Ishihama Y, Sato T, Nagasu T, Oda Y (2005) Specificity of immobilized metal affinity-based IMAC/C18 tip enrichment of phosphopeptides for protein phosphorylation analysis. Anal Chem 77:5144–5154
16. Yu Z, Han G, Sun S, Jiang X, Chen R, Wang F, Wu R, Ye M, Zou H (2009) Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. Anal Chim Acta 636:34–41
17. Zhou H, Ye M, Dong J, Han G, Jiang X, Wu R, Zou H (2008) Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. J Proteome Res 7:3957–3967
18. Zhou H, Low TY, Hennrich ML, van der Toorn H, Schwend T, Zou H, Mohammed S, Heck AJ (2011) Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. Mol Cell Proteomics 10(M110):006452
19. Jensen SS, Larsen MR (2007) The impact of different experimental procedures on phosphopeptide enrichment techniques. Accepted for Rapid Comm, Mass Spectrom
20. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. Mol Cell Proteomics 7:661–671
21. Kocher T, Allmaier G, Wilm M (2003) Nanoelectrospray-based detection and sequencing of substoichiometric amounts of phosphopeptides in complex mixtures. J Mass Spectrom 38:131–137
22. Gobom J, Nordhoff E, Mirgorodskaya E, Ekman R, Roepstorff P (1999) Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom 34:105–116
23. Engholm-Keller K, Birck P, Storling J, Pociot F, Mandrup-Poulsen T, Larsen MR (2012) TiSH–a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. J Proteomics 75:5749–5761
24. Liu S, Zhang C, Campbell JL, Zhang H, Yeung KK, Han VK, Lajoie GA (2005) Formation of phosphopeptide-metal ion complexes in liquid chromatography/electrospray mass spectrometry and their influence on phosphopeptide detection. Rapid Commun Mass Spectrom 19:2747–2756
25. Thingholm TE, Jensen ON (2009) Enrichment and characterization of phosphopeptides by immobilized metal affinity chromatography (IMAC) and mass spectrometry. Methods Mol Biol 527(47–56):xi

Chapter 9

The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides

Tine E. Thingholm and Martin R. Larsen

Abstract

Titanium dioxide (TiO_2) has very high affinity for phosphopeptides and in recent years it has become one of the most popular methods for phosphopeptide enrichment from complex biological samples. Peptide loading onto TiO_2 resin in a highly acidic environment in the presence of 2,5-dihydroxybenzoic acid (DHB), phthalic acid, lactic acid, or glycolic acid has been shown to improve selectivity significantly by reducing unspecific binding of non-phosphorylated peptides. The phosphopeptides bound to the TiO_2 are subsequently eluted from the chromatographic material using an alkaline buffer. TiO_2 chromatography is extremely tolerant towards most buffers used in biological experiments, highly robust and as such it has become the method of choice in large-scale phosphoproteomics. Here we describe a batch mode protocol for phosphopeptide enrichment using TiO_2 chromatographic material followed by desalting and concentration of the sample by reversed phase micro-columns prior to downstream MS and LC-MS/MS analysis.

Key words Protein phosphorylation, Phosphopeptide enrichment, Titanium dioxide chromatography, Mass spectrometry

1 Introduction

For some years the adsorption of proteins to titanium dioxide (TiO_2) films has been studied with the aim of determining a method for pursuing bioelectrochemical studies of protein functions [1]. More interestingly for phosphoproteomic studies, TiO_2 has been shown to have affinity for phosphate ions from aqueous solutions [2, 3] and recently, TiO_2 chromatography has been adapted as an efficient alternative to already existing methods for phosphopeptide enrichment from complex samples. In 2004 several groups introduced the enrichment of phosphopeptides using TiO_2 material in combination with mass spectrometric (MS) analysis [4–6]. For example, Pinkse and coworkers described the ability of TiO_2 to selectively bind phosphorylated peptides using an online two-dimensional liquid chromatography (LC) MS setup

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_9,

with spherical particles of TiO_2 (Titansphere) as the first dimension and reversed phase (RP) material as the second dimension [5]. The sample was loaded onto a TiO_2 column in acidic conditions (pH 2.9) to promote the binding of phosphopeptides to the TiO_2 particles. The unbound non-phosphorylated peptides were trapped on the RP column. After elution from the RP column the non-phosphorylated peptides were analyzed using nanoLC-ESI-MS/MS. The phosphopeptides were subsequently eluted from the TiO_2 column using an alkaline buffer (pH 9.0), concentrated on the RP pre-column and analyzed using nanoLC-ESI-MS/MS [5]. However, the buffers used for these initial TiO_2 enrichments all resulted in significant binding of non-phosphopeptides. In 2005, we introduced an offline setup for TiO_2 chromatography in which much stronger buffer conditions were applied and included the use of 2,5-dihydroxybenzoic acid (DHB) and high concentration of trifluoroacetic acid (TFA) in the loading buffer, which significantly reduced unspecific binding from non-phosphorylated peptides [7]. In addition, ammonia solution at even higher pH (pH 11.3) was shown to elute phosphorylated peptides from the TiO_2 column more efficiently than pH 9 and thereby improve phosphopeptide recovery [7]. Due to problems associated with the use of high amount of DHB, such as polymerization and contamination of the LC system we investigated other multifunctional acids such as phthalic acid [8] and glycolic acid [9] and found the latter to be ideal for large-scale phosphoproteomic studies.

The high selectivity of TiO_2 towards phosphorylated peptides makes it a powerful tool for phosphoproteomic studies, also when used in combination with other phosphopeptide enrichment methods such as Immobilized Metal Affinity Chromatography (IMAC) (*see* Chapter 8) termed Sequential elution from IMAC (SIMAC) [10] (*see* Chapter 10) or TiSH [11] (*see* Chapter 11) methods. In addition, TiO_2 chromatography of phosphorylated peptides is extremely tolerant towards most buffers and salts used in biochemistry and cell biology laboratories [9]. The offline setup is simple and fast, and does not require expensive equipment.

2 Materials

2.1 Model Proteins

1. Transferrin (human) was a gift from ACE Biosciences A/S. Serum albumin (bovine), α-lactoglobulin (bovine), carbonic anhydrase (bovine), α-casein (bovine), β-casein (bovine), ovalbumin (chicken), ribonuclease B (bovine), alcohol dehydrogenase (Baker's yeast), myoglobin (whale skeletal muscle), lysozyme (chicken), α-amylase (*Bacillus* species) were from Sigma (St. Louis. MO, USA). Of these 12 proteins 3 are phosphorylated: α-casein, β-casein, and ovalbumin, yielding about 15 phosphorylated peptides depending on the purity of the batches.

2.2 Reduction, Alkylation, and Digestion of Model Proteins

1. Triethylammonium bicarbonate.
2. Dithiothreitol (DTT).
3. Iodoacetamide.
4. Modified trypsin.
5. Acetone.

2.3 Cell Lysis and Titanium Dioxide (TiO_2) Chromatography

1. Titanium dioxide (TiO_2) beads (Titansphere, 5 μm, GL sciences Inc.).
2. Low-binding microcentrifuge tubes 1.7 mL.
3. Acetonitrile, HPLC Grade.
4. TiO_2 Loading Buffer: 1 M glycolic acid in 5 % trifluoroacetic acid (TFA), 80 % acetonitrile.
5. TiO_2 Washing Buffer 1: 1 % TFA, 80 % acetonitrile.
6. TiO_2 Washing Buffer 2: 0.1 % TFA, 10 % acetonitrile.
7. TiO_2 Elution Buffer: 1 % ammonia water (40 μL ammonia Solution (25 %) in 960 μL UHQ water).
8. Formic acid.
9. Milli-Q water (UHQ water) (*see* **Notes 1** and **2**).
10. PhosStop.

2.4 Reversed Phase (RP) Micro-columns

1. POROS Oligo R3 reversed phase material (PerSeptive Biosystems, Framingham, MA, USA).
2. GELoader tips (Eppendorf, Hamburg, Germany) or p200 pipette tips depending on the size of the column needed.
3. 3 M Empore C18 disk (3 M, Bioanalytical Technologies, St. Paul, MN, USA).
4. 1–5 mL disposable syringes fitted to GeLoader tip or p200 tips by using a pipette tip cut in both ends.
5. RP Washing Buffer: 0.1 % TFA.
6. RP Elution Buffer (for LC-ESI MS/MS analysis): 70 % acetonitrile, 0.1 % TFA.
7. 2,5-dihydroxybenzoic acid (DHB) Elution Buffer (for MALDI MS analysis): 20 mg/mL DHB in 50 % acetonitrile, 1 % ortho-phosphoric acid.

2.5 Other Materials

1. Tabletop centrifuge.
2. pH meter.
3. Thermomixer.
4. Shaker.
5. Vacuum centrifuge.

2.6 Analysis by Mass Spectrometry

1. Mass spectrometer capable of performing MS/MS—preferentially a high-resolution/high mass accuracy instrument (Q-TOFs (Waters, ABSciex, Bruker, and Agilent) or Orbitrap based mass spectrometer (Thermo Fisher Scientific)) interfaced to a nanoHPLC (e.g., Dionex 3000 ultimate LC system (Thermo Fisher Scientific)) with a 50–100 μm i.d. RP capillary column setup for highly sensitive online peptide separation can be used. For simpler samples a MALDI MS instrument can be used (e.g., Bruker Ultraflex (Bruker Daltonics, Bremen, Germany)).
2. Software for processing of raw mass spectrometry data files and generation of peak lists for searching against a protein database (e.g., Uniprot) Analysis software such as Mascot/Mascot Distiller (Matrix Science, London, UK) (data from most vendors and instruments), Proteome Discoverer (Thermo Scientific, Bremen, Germany) (data from Thermo instruments), MaxQuant [15] (high resolution data from Thermo Orbitrap instruments and certain Bruker and ABSciex Q-TOFs) and the TransProteomicPipeline [16] (vendor independent).

3 Methods

The principle of the TiO_2 method described in this chapter is illustrated by the purification of phosphopeptides from a peptide mixture originating from tryptic digestions of 12 standard proteins (Model proteins) and from 150 μg of tryptic peptides from a HeLa cell lysate (*see* **Notes 3** and **4**).

The TiO_2 purification method is simple and straightforward. It is fast and efficient for enrichment of phosphopeptides even from highly complex samples (e.g., [11, 12]) (*see* **Note 5**). The experimental setup of the method is illustrated in Fig. 1. For illustrating the anticipated results, first a peptide mixture originating from tryptic digestions of 12 standard proteins was subjected to TiO_2. Secondly, the TiO_2 procedure was applied to 150 μg of tryptic peptides derived from a HeLa cell lysate.

3.1 Digestion of Model Proteins and the HeLa Cell Lysate

1. Dissolve each protein in 50 mM triethylammonium bicarbonate (TEAB), pH 7.8, 10 mM DTT and incubate at 37 °C for 1 h. After reduction, add 20 mM iodoacetamide and incubate the samples at room temperature for 1 h in the dark.
2. Digest each protein using trypsin (1–2 % w/w) at 37 °C for 12 h.
3. Lyse HeLa cells in 6 M urea, 2 M thiourea containing phosphatase inhibitors. Precipitate the proteins using 10 volume excess of ice-cold acetone and incubate overnight at −20 °C. Centrifuge the sample at 14,000 × *g* and wash the pellet twice with ice-cold acetone. Redissolve the pellet in 50 μL 6 M urea,

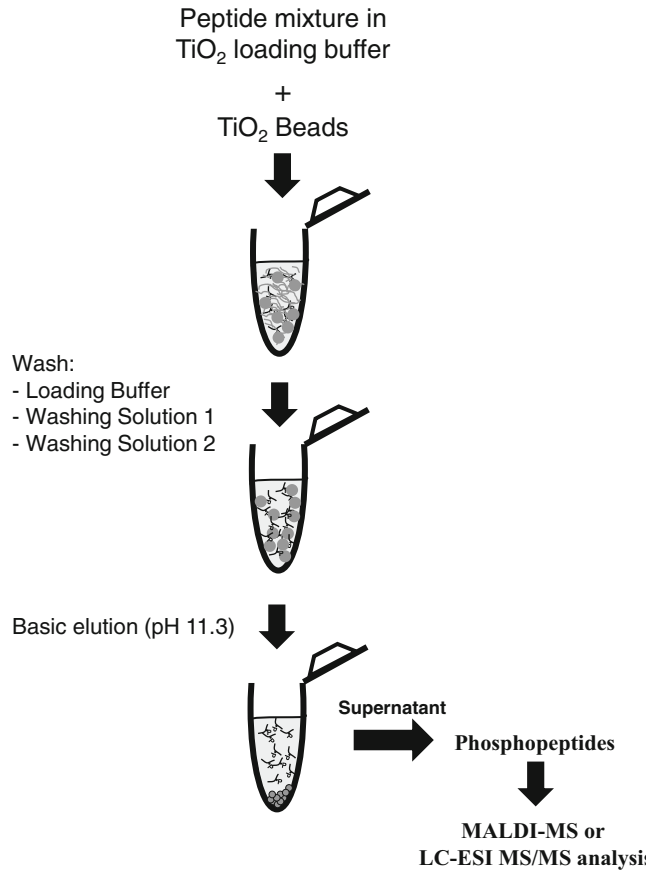


Fig. 1 The strategy used for the enrichment of phosphopeptides using TiO_2 material. The peptide sample is diluted with the TiO_2 Buffer and TiO_2 beads are added. The sample is incubated for 5–10 min. The TiO_2 beads are washed with washing solutions and finally the phosphopeptides are eluted from the TiO_2 material using basic elution conditions

2 M thiourea, 10 mM DTT containing 1 μg endoproteinase Lys-C and incubate at room temperature for 2 h. After incubation dilute the sample 10× with 50 mM TEAB, pH 7.8 containing 20 mM iodoacetamide and incubate for 1 h in the dark at room temperature. After incubation, add trypsin (1–2 % w/w) and incubate the sample overnight at room temperature.

3.2 Batch Mode TiO_2 Phosphopeptide Purification

1. Add acetonitrile, TFA, and glycolic acid to your peptide mixture to obtain TiO_2 Loading Buffer conditions (80 % acetonitrile, 5 % TFA, and 1 M glycolic acid) (*see* **Note 6**) or dilute the sample at least 10× with the TiO_2 Loading Buffer.
2. Add 0.6 mg TiO_2 beads per 100 μg peptide solution (*see* **Note 7**).

3. Place the tubes on a shaker (highest shaking) at room temperature for 5–10 min.
4. After incubation centrifuge to pellet the beads (table centrifuge <15 s).
5. Transfer the supernatant to another low-binding tube and incubate it for a second round of incubation with 0.3 mg TiO_2 beads. This step may be repeated more times in order to recover as many phosphopeptides as possible.
6. Pool the TiO_2 beads from the 2–3 incubations using 100 µL Loading Buffer and transfer them to a new low-binding microcentrifuge tube (*see* **Note 8**).
7. Vortex the solution for 10 s and then centrifuge to pellet the beads. Remove the supernatant.
8. Wash the beads with 70–100 µL (*see* **Note 9**) Washing Buffer 1—mix for 10 s and then centrifuge to pellet the beads.
9. Wash the beads with 70–100 µL Washing Buffer 2—mix for 10 s and then centrifuge to pellet the beads. This step is important to remove peptides that bind in a HILIC mode to TiO_2 (*see* **Note 10**).
10. Dry the beads for 5–10 min in the vacuum centrifuge or on the table for 30 min.
11. Elute the phosphopeptides with 100–200 µL Elution Buffer—mix well and leave the solution on a shaker for 10 min to allow an efficient elution.
12. Centrifuge the solution for 1 min and pass the supernatant (TiO_2-eluate) over a small stage tip filter [13] (C8 stage tip) into a new low-binding tube to recover the liquid without TiO_2 beads.
13. Wash the beads with 30 µL Elution Buffer and pool the wash with the TiO_2-eluate from the previous step.
14. Elute potential bound peptides from the C8 filter with 5 µL 30 % acetonitrile and pool with the TiO_2-eluate from **steps 12** to **13**.
15. Lyophilize the eluted peptides (*see* **Note 11**) or acidify the eluate with 1 µL Formic Acid per 10 µL eluate for direct cleanup of the phosphopeptides using RP material prior to downstream analyses (e.g., HILIC fractionation [11]).

3.3 Poros Oligo R3 Reversed Phase (RP) Micro-column Desalting/ Concentration of the Sample

Use RP GELoader tip micro-columns of ~6–10 mm or p200 pipette tips micro-columns (1–2 cm) depending on the amount of material to be purified. Here, it is illustrated for the p200 pipette tip.

1. Suspend Poros Oligo R3 reversed phase (RP) material in 200 µL 100 % acetonitrile.

2. Prepare a p200 pipette tip micro-column by stamping out a small plug of C_{18} material from a 3 M Empore™ C_{18} extraction disk and place it in the constricted end of the tip.
3. Pack Poros Oligo R3 RP beads on top of the p200 stage tip until the size of the column is 1–2 cm.
4. Load the acidified phosphopeptide sample slowly onto the RP micro-column (~1 drop/s).
5. Wash the RP micro-column using 60 µL RP Washing Buffer.
6. Elute the phosphopeptides from the RP micro-column using 40–60 µL RP Elution Buffer, followed by lyophilization of the phosphopeptides (N.B. For MALDI MS analysis the peptides can be eluted off the GeLoader tip RP micro-column directly onto the MALDI target using 1 µL DHB solution. After crystallization the sample is ready for MALDI MS analysis).
7. Redissolve the lyophilized phosphopeptides in 0.5 µL 100 % formic acid and dilute immediately to 10 µL with UHQ water. The sample is then ready for LC-ESI-MS^n analysis.

3.4 µHPLC Tandem Mass Spectrometry (LC-MS/MS) Analysis

For LC-MS/MS analysis of purified phosphopeptides a standard strategy as described below can be used. A typical nanoLC setup would include a 0.075 mm × 200 mm analytical column packed with 3 µm RP resin interfaced with a high resolution/mass accuracy mass spectrometer as described in our original paper [11]. The number of phosphopeptides identified in the analysis can be increased by maximizing the resolution of the nanoLC separation via longer columns (e.g., 50 cm) and smaller chromatographic particle sizes (e.g., 1.9 µm). Alternatively, a two column system can be utilized using a 0.1 mm × 20 mm pre-column packed with RP resin (3–5 µm) combined with an analytical column as described above. A two column system is described below.

1. The phosphopeptides are redissolved in 0.1 % TFA and loaded onto a pre-column as described above using a µHPLC system (e.g., Dionex or EASY-LC) at a loading speed of 5 µL/min.
2. The phosphopeptides are eluted directly onto the analytical column (e.g., 0.075 mm × 200 mm) using a gradient (60–120 min) from 0 to 35 % B-Buffer (e.g., A-Buffer: 0.1 % formic acid; B-Buffer: 90 % acetonitrile, 0.1 % TFA) at an elution speed of 2–300 nL/min.
3. The phosphopeptides are eluted directly into a tandem mass spectrometer and analyzed by Data Dependent Analysis.

LC-ESI-MS^n analysis of multi-phosphorylated peptides can be improved by redissolving the phosphopeptides by sonication in an EDTA containing buffer prior to LC-ESI-MS^n analysis [14] (*see* **Note 12**).

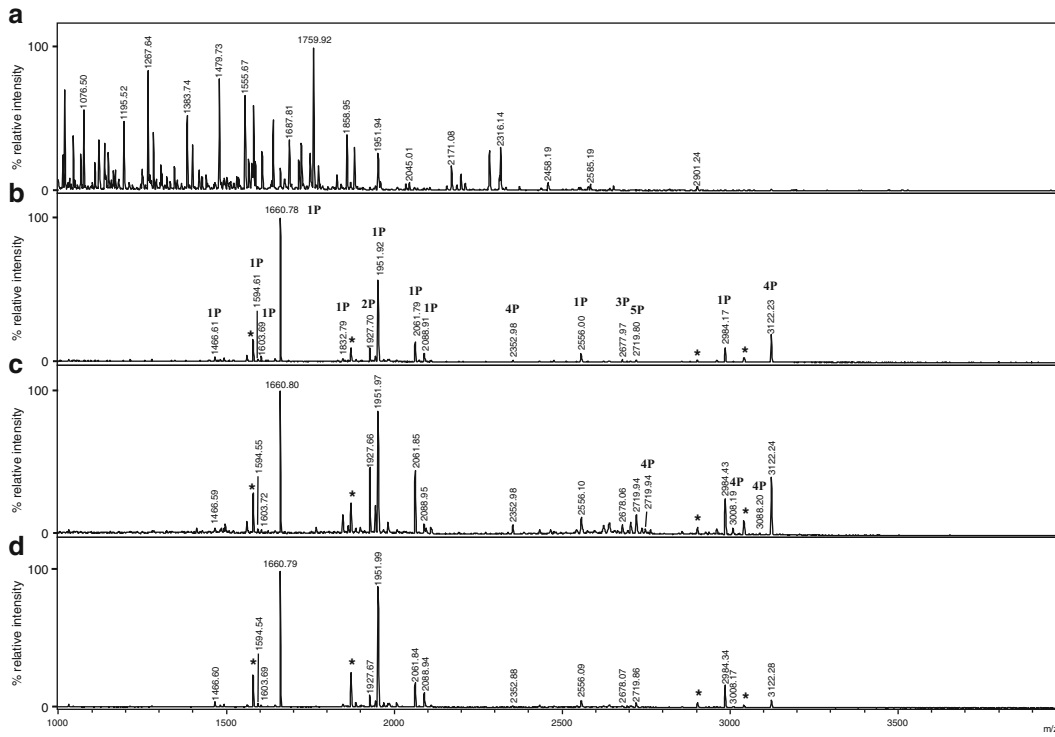


Fig. 2 Results obtained from 1 pmol peptide mixture using the TiO_2 strategy. (**a**) MALDI MS peptide mass map of the direct analysis of the tryptic peptides. (**b**) MALDI MS peptide mass map of phosphopeptides purified by TiO_2 from the standard peptides mixture dissolved in 0.1 % TFA. (**c**) MALDI MS peptide mass map of phosphopeptides purified by TiO_2 from the standard peptides mixture dissolved in 5 % SDS and 25 mM EDTA. (**d**) MALDI MS peptide mass map of phosphopeptides purified by TiO_2 from the standard peptides mixture dissolved in 1 M KCl and 25 mM EDTA. The number of phosphate groups on the individual phosphopeptides is indicated by "#P". *Asterisk* indicates the metastable loss of phosphoric acid

An example of the results obtained by TiO_2 chromatography is shown in Fig. 2. The figure shows the MALDI MS results obtained on a Bruker Ultraflex from a dried droplet sample preparation of 0.25 pmol peptide mixture without phosphopeptide enrichment (Fig. 2a). The MALDI MS peptide mass map of the TiO_2 enriched phosphopeptides from 0.25 pmol of peptide mixture is shown in (Fig. 2b). The unique tolerance towards biological buffers is illustrated by the enrichment of phosphopeptides from the peptide mixture in the presence of 50 mM EDTA which will normally strip the iron from the chromatographic material used in IMAC. The MALDI MS peptide mass map is shown in (Fig. 2c). The phosphopeptides are illustrated in (Fig. 2c) by asterisks (*see* **Note 13**).

An example of the results obtained using the present protocol for enrichment of phosphopeptides from a total of 150 μg peptides derived by tryptic digestion from a HeLa cell lysate is shown in Fig. 3. The enriched phosphopeptides were separated on a Dionex

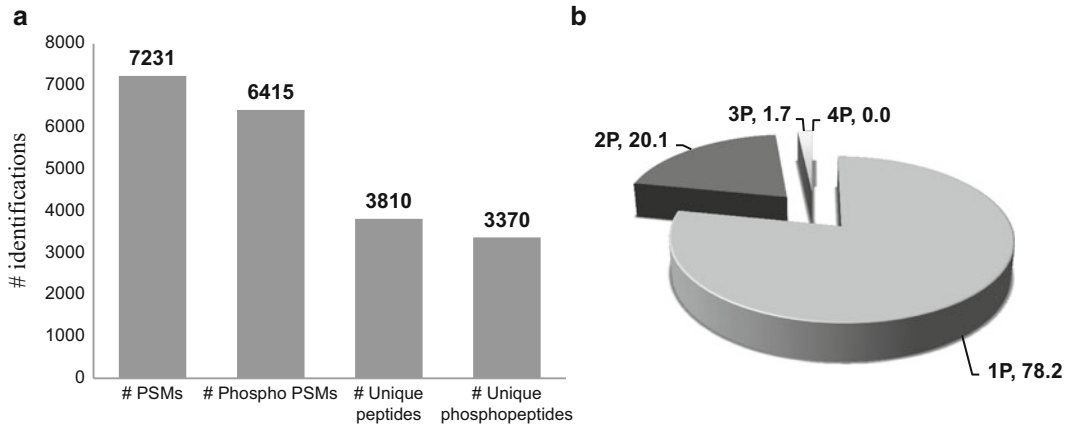


Fig. 3 Results obtained from the enrichment of phosphorylated peptides from acetone precipitated proteins from HeLa cells using TiO_2 chromatography. (**a**) Overview of the number of peptide spectral matches (PSMs) and unique phosphopeptides identified in the TiO_2 enriched fraction. (**b**) Percentage distribution of the number of phosphate groups on the phosphopeptides identified in the TiO_2 enriched fraction

3000 ultimate LC system using homemade RP capillary columns (20–25 cm) directly into a Q-Exactive Plus MS instrument from Thermo Fisher Scientific. The peptides were separated using a 90 min gradient from 0 to 25 % B Buffer (90 % acetonitrile in 0.1 % formic acid). The MS instrument was set to isolate and fragment 12 parent ions per MS cycle (MS and MS/MS resolution was set to 70,000 and 35,000 at 200 *m/z*, respectively; MS and MS/MS AGC target was 1E6 and 5E4, respectively; normalized collision energy was 30; isolation window was 1.5 Da). Here a total of 3370 unique phosphopeptides were identified from the 150 μg of starting material, using the Proteome Discoverer 1.4.1.14 (SwissProt_2014_04 (20,340 entries)) with an enrichment percentage of about 88 % phosphopeptides (*see* Fig. 3a). Of these, 2634 were singly phosphorylated and 735 had two or more phosphate groups (*see* Fig. 3b).

4 Notes

1. It is important to obtain the highest purity of all chemicals used.
2. All solutions should be prepared in UHQ water.
3. Always start by testing the method using a model peptide mixture. It is important to freshly prepare the peptide mixture as peptides bind to the surface of the plastic tubes in which they are stored. In addition, avoid transferring the peptide sample between different tubes to minimize adsorptive losses of the sample to the plastic surfaces.

4. The peptide mixtures used for the experiments illustrated in this chapter contained peptides originating from tryptic digestions of 1 pmol of each of 12 standard proteins and of 150 μg of a HeLa cell lysate. Experiments have shown that the presented method is sensitive down to the low femtomole level [8].
5. Frequently, the TiO_2 enrichment method described here is combined with either peptide pre- or post-fractionation methods such as strong cation exchange (SCX) (e.g., [15, 16]) or Hydrophilic Interaction Liquid Chromatography (HILIC) (e.g., [11, 12, 17]) in order to increase the phosphopeptide coverage in large-scale phosphoproteomic studies.
6. If you have 100 μL peptide sample, you can add 50 μL water, 50 μL 100 % TFA, 800 μL acetonitrile, and 76 mg glycolic acid to obtain the proper TiO_2 Loading Buffer conditions.
7. The optimal amount of TiO_2 beads to use in order to reduce nonspecific binding and optimize phosphopeptide yield is 0.6 mg TiO_2 per 100 μg of peptide starting material (*see* [11] for further information). This will of course change depending on the source of biological material used as TiO_2 selectively enriches other biomolecules (reviewed in [18]) such as sialylated glycopeptides [19] and acidic lipids [20] commonly found in membrane fractions.
8. The transfer to a new tube is performed due to the fact that peptides stick to plastic and can be eluted from the plastic surface in the last elution step resulting in contamination with non-modified peptides.
9. For large-scale analysis, where more TiO_2 beads are used, larger volumes of the buffers should be used.
10. TiO_2 is an efficient HILIC material and hydrophilic peptides can bind to the material when loaded in high organic solvent. The inclusion of 5 % TFA and 1 M glycolic acid should prevent most hydrophilic non-modified peptides from binding; however, some can still be found in the elution from TiO_2. Therefore in order to eliminate any binding from non-phosphorylated hydrophilic peptides this last Washing Buffer is important. For membrane preparations the last washing supernatant will contain neutral glycopeptides which can then be analyzed further.
11. If ammonia is left in the solution, it will interfere with deglycosylation procedures (membrane fractions) or a subsequent SIMAC purification [11].
12. Titanium dioxide was originally thought to have a preference for mono-phosphorylated peptides likely due to the fact that it is difficult to elute multi-phosphorylated peptides from the

TiO_2 micro-column once they have bound [9], as performed in the original DHB TiO_2 procedure [7]. However, since the procedure is now performed in batch mode the elution takes longer time and plenty of multi-phosphorylated peptides can be eluted with the TiO_2 Elution Buffer described here [11]. The combination of sequential elution from Immobilized Metal Affinity Chromatography (IMAC) and TiO_2 has made it possible to enrich for mono- as well as multi-phosphorylated peptides as described in the SIMAC protocol in Chapter 10.

13. The results obtained using this protocol will differ according to the mass spectrometer used for the analysis of the phosphopeptides, not only between MALDI MS and ESI MS but also within different MALDI MS instruments, depending on laser optics, laser frequency, instrumental configuration, sensitivity, etc.

Acknowledgements

This work was supported by the Danish Natural Science and Medical Research Councils (grant no. 10-082195 (T.E.T)) and the Lundbeck Foundation (M.R.L—Junior Group Leader Fellowship).

References

1. Topoglidis E, Cass AEG, Gilardi G, Sadeghi S, Beaumont N, Durrant JR (1998) Protein adsorption on nanocrystalline TiO2 films: an immobilization strategy for bioanalytical devices. Anal Chem 70:5111–5113
2. Connor PA, Dobson KD, McQuillan J (1999) Infrared spectroscopy of the TiO2/aqueous solution interface. Langmuir 15:2402–2408
3. Connor PA, Mcquillan AJ (1999) Phosphate adsorption onto TiO2 from aqueous solutions: an in situ internal reflection infrared spectroscopic study. Langmuir 15:2916–2921
4. Sano A, Nakamura H (2004) Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. Anal Sci 20:565
5. Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76:3935–3943
6. Kuroda I, Shintani Y, Motokawa M, Abe S, Furuno M (2004) Phosphopeptide-selective column-switching RP-HPLC with a titania precolumn. Anal Sci 20:1313–1319
7. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4:873–886
8. Thingholm TE, Jorgensen TJ, Jensen ON, Larsen MR (2006) Highly selective enrichment of phosphorylated peptides using titanium dioxide. Nat Protoc 1:1929–1935
9. Jensen SS, Larsen MR (2007) Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. Rapid Comm Mass Spectrom 21:3635
10. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC - a phosphoproteomic strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. Mol Cell Proteom 7(4):661–671
11. Engholm-Keller K, Birck P, Storling J, Pociot F, Mandrup-Poulsen T, Larsen MR (2012) TiSH--a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. J Proteomics 75:5749–5761

12. Engholm-Keller K, Hansen TA, Palmisano G, Larsen MR (2011) Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO(2) chromatography applied to proximal EGF signaling. J Proteome Res 10:5383–5397
13. Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75:663–670
14. Liu S, Zhang C, Campbell JL, Zhang H, Yeung KK, Han VK, Lajoie GA (2005) Formation of phosphopeptide-metal ion complexes in liquid chromatography/electrospray mass spectrometry and their influence on phosphopeptide detection. Rapid Commun Mass Spectrom 19:2747–2756
15. Pinkse MW, Mohammed S, Gouw JW, van Breukelen B, Vos HR, Heck AJ (2008) Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. J Proteome Res 7:687–697
16. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648
17. McNulty DE, Annan RS (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Mol Cell Proteomics 7:971–980
18. Engholm-Keller K, Larsen MR (2011) Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds--applications in acidic modification-specific proteomics. J Proteomics 75:317–328
19. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry. Mol Cell Proteomics 6:1778–1787
20. Calvano CD, Jensen ON, Zambonin CG (2009) Selective extraction of phospholipids from dairy products by micro-solid phase extraction based on titanium dioxide microcolumns followed by MALDI-TOF-MS analysis. Anal Bioanal Chem 394:1453–1461

Chapter 10

Sequential Elution from *IMAC* (SIMAC): An Efficient Method for Enrichment and Separation of Mono- and Multi-phosphorylated Peptides

Tine E. Thingholm and Martin R. Larsen

Abstract

Phosphoproteomics relies on methods for efficient purification and sequencing of phosphopeptides from highly complex biological systems, especially when using low amounts of starting material. Current methods for phosphopeptide enrichment, e.g., Immobilized Metal ion Affinity Chromatography and titanium dioxide chromatography provide varying degrees of selectivity and specificity for phosphopeptide enrichment. The number of multi-phosphorylated peptides identified in most published studies is rather low. Here we describe a protocol for a strategy that separates mono-phosphorylated peptides from multiply phosphorylated peptides using Sequential elution from Immobilized Metal ion Affinity Chromatography. The method relies on the initial enrichment and separation of mono- and multi-phosphorylated peptides using Immobilized Metal ion Affinity Chromatography and a subsequent enrichment of the mono-phosphorylated peptides using titanium dioxide chromatography. The two separate phosphopeptide fractions are then subsequently analyzed by mass spectrometric methods optimized for mono-phosphorylated and multi-phosphorylated peptides, respectively, resulting in improved identification of especially multi-phosphorylated peptides from a minimum amount of starting material.

Key words Phosphopeptide enrichment, Multi-phosphorylated peptides, Immobilized metal affinity chromatography, Sequential elution, Titanium dioxide chromatography, Mass spectrometry

1 Introduction

Several techniques exist for phosphopeptide enrichment prior to mass spectrometric analysis. Today the most commonly used methods are Immobilized Metal Affinity Chromatography (IMAC) [1–3] and titanium dioxide (TiO_2) chromatography [4–7] (*see* Chapters 8 and 9). Recent studies comparing three different phosphopeptide enrichment methods including phosphoramidate chemistry (PAC) [8], IMAC and TiO_2 chromatography showed that each method isolated distinct, partially overlapping segments of a phosphoproteome, whereas none of the tested methods was

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_10,

able to provide a whole phosphoproteome [9]. This is in itself not surprising as the three different methods apply completely different chemistries for phosphopeptide capture, numerous protocols for IMAC and TiO_2 exist and the purification efficiency can be very variable for both IMAC and TiO_2 depending on the person who is performing the analysis.

One of the challenges in large-scale phosphoproteomics is the analysis of multi-phosphorylated peptides. Multi-phosphorylated peptides are in general suppressed in the ionization process in the mass spectrometric (MS) analysis in the presence of mono- or non-phosphorylated peptides and therefore the chance to detect them by tandem MS (MS/MS) analysis is limited. In addition, most mass spectrometers are only able to perform a limited number of MS/MS in a given time period resulting in the negligence of the less abundant multi-phosphorylated peptides. Furthermore, in collision induced dissociation (CID) the major fragmentation pathway is the loss of phosphoric acid usually resulting in poor peptide backbone fragmentation. Consequently, little sequence information and lower identification rates are obtained. This is especially evident for multi-phosphorylated peptides which lose more phosphoric acid molecules. Several other kind of fragmentation methods exist which can increase the identification of multi-phosphorylated peptides. Optimized phosphorylation-directed multistage tandem MS ($pdMS^3$) [10, 11], multistage activation (MSA) [12], higher energy collision dissociation (HCD) [13] or Electron capture/transfer dissociation (ECD/ETD) [14, 15] could provide better identification for multi-phosphorylated peptides. However, in order to set up the special experimental parameters optimal for analysis of multi-phosphorylated peptides, such as normalized collision energy, fragmentation time and number of ions used for fragmentation, the multi-phosphorylated peptides have to be separated from the mono-phosphorylated peptides prior to LC-MS/MS analysis.

Previously, we developed a method for separation of mono-phosphorylated peptides from multiply phosphorylated peptides where we are using *S*equential elution from *IMAC* (SIMAC) [11]. In this strategy the peptide mixture is incubated with IMAC beads, which have a stronger selectivity for multi-phosphorylated peptides than for mono-phosphorylated peptides [16]. After incubation, the sample is split in three "elution" fractions (*see* Fig. 1); an IMAC flow-through fraction, an acidic (1 % TFA) fraction and a basic (pH 11.3) fraction. The IMAC flow-through and acidic fractions which contain predominantly mono-phosphorylated and a significant number of non-phosphorylated peptides are further submitted to TiO_2 chromatography to achieve pure phosphorylated fractions prior to tandem MS analysis. Alternatively, the two fractions can be pooled prior to TiO_2 enrichment. The basic fraction is analyzed directly by MS/MS analysis without further TiO_2 purification, as this sample in general is relative free of non-phosphorylated peptides.

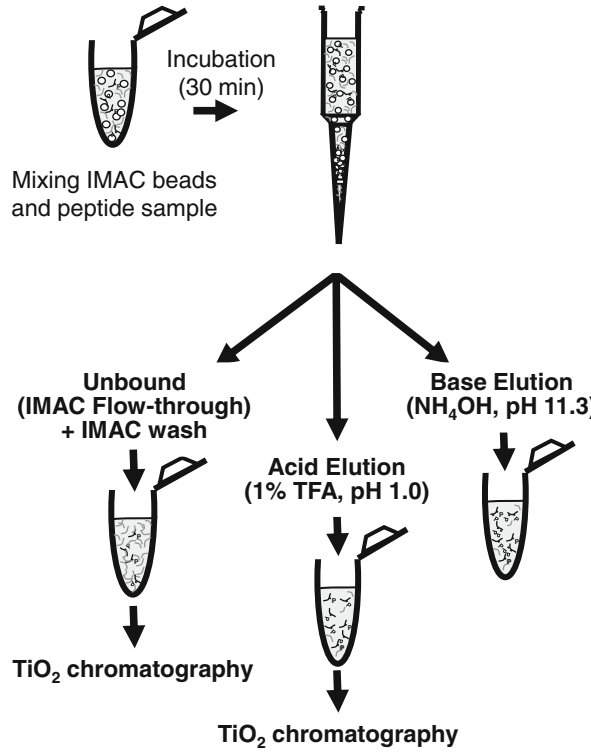


Fig. 1 The SIMAC strategy used for the enrichment and separation of mono- from multi-phosphorylated peptides. The peptide sample is mixed with the IMAC beads and incubated for 30 min in a Thermomixer at room temperature. After incubation, the beads are packed into a GELoader tip forming an IMAC micro-column. The IMAC flow-through is collected and further enriched using TiO_2 chromatography. The mono-phosphorylated peptides are eluted from the IMAC micro-column using acidic elution conditions (1 % TFA, pH 1.0) and for complex samples this eluate is also further enriched using TiO_2 chromatography or combined with the IMAC-FT prior to TiO_2 enrichment. The multi-phosphorylated peptides are subsequently eluted from the IMAC micro-column using basic elution conditions (ammonia water, pH 11.3). The Figure is taken from [25]

SIMAC greatly improves the number of phosphorylation sites identified even from very low amounts of starting material and offers a way to identify and characterize multi-phosphorylated peptides at large-scale levels [11] (*see* also Chapter 11).

2 Materials

2.1 Model Proteins

1. Transferrin (human) was a gift from ACE Biosciences A/S. Serum albumin (bovine), beta-lactoglobulin (bovine), carbonic anhydrase (bovine), beta-casein (bovine), alpha-casein (bovine), ovalbumin (chicken), ribonuclease B (bovine

pancreas), alcohol dehydrogenase (Baker yeast), myoglobin (whale skeletal muscle), lysozyme (chicken), and alpha-amylase (bacillus species) were from Sigma (St. Louis. MO, USA).

2.2 Reduction, Alkylation, and Digestion of Proteins

1. Triethylammonium bicarbonate.
2. Dithiothreitol (DTT).
3. Iodoacetamide.
4. Modified trypsin.
5. Acetone.

2.3 Reduction, Alkylation, and Digestion of HeLa Proteins

1. Lysis Buffer: 6 M urea, 2 M thiourea, 1× PhosSTOP phosphatase inhibitors.
2. Dithiothreitol (DTT).
3. Iodoacetamide.
4. Endoproteinase Lys-C.
5. Triethylammonium bicarbonate.
6. Modified trypsin.
7. PhosStop.

2.4 Immobilized Metal ion Affinity Chromatography (IMAC)

1. Iron-coated PHOS-select™ metal chelate beads (Sigma®), stored at −20 °C.
2. IMAC Loading Buffer: 0.1 % trifluoroacetic acid (TFA), Protein Sequencer Grade, 50 % acetonitrile, HPLC Grade.
3. GELoader tips (Eppendorf (20 µL) or Bio-Rad (200 µL)).
4. Low-binding microcentrifuge tubes 1.7 mL.
5. 1–5 mL disposable syringes fitted to GeLoader tip or p200 tips by using a pipette tip cut in both ends.
6. IMAC Elution Buffer 1: 1 % TFA, 20 % acetonitrile.
7. IMAC Elution Buffer 2: 1 % ammonia water (40 µL ammonia solution (25 %), 980 µL UHQ water (pH ~ 11)), make fresh as required.
8. Formic acid.
9. Milli-Q water (UHQ water) (*see* **Notes 1** and **2**).

2.5 Titanium Dioxide (TiO_2) Chromatography

1. Titanium dioxide (TiO_2) beads (Titansphere, 5 µm, GL sciences Inc.).
2. Low-binding microcentrifuge tubes 1.7 mL.
3. 3 M Empore C8 disk (3 M, Bioanalytical Technologies, St. Paul, MN, USA).
4. Acetonitrile, HPLC Grade.
5. TiO_2 Loading Buffer: 1 M glycolic acid in 5 % trifluoroacetic acid (TFA), 80 % acetonitrile.

6. TiO_2 Washing Buffer 1: 1 % TFA, 80 % acetonitrile.
7. TiO_2 Washing Buffer 2: 0.1 % TFA, 10 % acetonitrile.
8. TiO_2 Elution Buffer: 1 % ammonia water (40 μL ammonia solution (25 %) in 960 μL UHQ water).
9. Formic acid.

2.6 Reversed Phase (RP) Micro-columns

1. POROS Oligo R3 reversed phase material (PerSeptive Biosystems, Framingham, MA, USA).
2. GELoader tips (Eppendorf, Hamburg, Germany) or p200 pipette tips depending on the size of the column needed.
3. 3 M Empore C18 disk (3 M, Bioanalytical Technologies, St. Paul, MN, USA).
4. 1–5 mL disposable syringes fitted to GeLoader tip or p200 tips by using a pipette tip cut in both ends.
5. RP Washing Buffer: 0.1 % TFA.
6. RP Elution Buffer (for LC-ESI MS/MS analysis): 70 % acetonitrile, 0.1 % TFA.
7. 2,5-dihydroxybenzoic acid (DHB) Elution Buffer (for MALDI MS analysis): 20 mg/mL DHB in 50 % acetonitrile, 1 % orthophosphoric acid.

2.7 Other Materials

1. Tabletop centrifuge.
2. pH meter.
3. Thermomixer.
4. Shaker.
5. Vacuum centrifuge.

2.8 Analysis by Mass Spectrometry

1. Mass spectrometer capable of performing MS/MS—preferentially a high-resolution/high mass accuracy instrument (Q-TOFs (Waters, ABSciex, Bruker, and Agilent) or Orbitrap based mass spectrometer (Thermo Fisher Scientific)) interfaced to a nanoHPLC (e.g., Dionex 3000 ultimate LC system (Thermo Fisher Scientific)) with a 50–100 μm i.d. RP capillary column setup for highly sensitive online peptide separation can be used. For simpler samples a MALDI MS instrument can be used (e.g., Bruker Ultraflex (Bruker Daltonics, Bremen, Germany)).
2. Software for processing of raw mass spectrometry data files and generation of peak lists for searching against a protein database (e.g., Uniprot) Analysis software such as Mascot/Mascot Distiller (Matrix Science, London, UK) (data from most vendors and instruments), Proteome Discoverer (Thermo Scientific, Bremen, Germany) (data from Thermo instruments), MaxQuant [15] (high resolution data from Thermo Orbitrap instruments and certain Bruker and ABSciex Q-TOFs) and the TransProteomicPipeline [16] (vendor independent).

3 Methods

The principle of the SIMAC method is illustrated in this chapter firstly using a peptide mixture originating from tryptic digestions of 12 standard proteins (Model proteins) (*see* **Notes 3** and **4**). The protocol is then applied to enrich for phosphorylated peptides from whole cell lysates from 150 µg of proteins from HeLa cells.

The SIMAC purification method is a simple and very straightforward method. It is fast and efficient for enrichment of phosphopeptides from even highly complex samples [17, 18]. The experimental setup of the method is illustrated in Fig. 1.

3.1 Digestion of Model Proteins and the HeLa Cell Lysate

1. Dissolve each protein in 50 mM triethylammonium bicarbonate (TEAB), pH 7.8, 10 mM DTT and incubate at 37 °C at 1 h. After reduction, add 20 mM iodoacetamide and incubate the samples at room temperature for 1 h in the dark.
2. Digest each protein using trypsin (1–2 % w/w) at 37 °C for 12 h.
3. Lyse HeLa cells in 6 M Urea, 2 M ThioUrea containing phosphatase inhibitors (PhosStop). Precipitate proteins using 10 volume excess of ice-cold acetone and incubate over night at −20 °C. Centrifuge the sample at 14,000 × *g* and wash the pellet twice with ice-cold acetone. Redissolve the pellet in 50 µL 6 M urea, 2 M thiourea, 10 mM DTT containing 1 µg endoproteinase Lys-C and incubate at room temperature for 2 h. After incubation, dilute the sample 10× with 50 mM TEAB, pH 7.8 containing 20 mM iodoacetamide and incubate for 1 h in the dark at room temperature. After incubation, add trypsin (1–2 % w/w) and place the sample at room temperature overnight.

3.2 Batch Mode Sequential Enrichment and Separation with IMAC Beads

Always adjust the amount of IMAC beads to the amount of sample in order to reduce the level of nonspecific binding from non-phosphorylated peptides. For 1 pmol tryptic digest use 7 µL IMAC beads (*see* Chapter 8). For more complex samples where more material is available, more IMAC beads should be used. This section is describing a protocol for using 150 µg tryptic digest from HeLa cells.

1. Transfer 50 µL IMAC beads to a fresh low-binding microcentrifuge tube 1.7 mL.
2. Wash the IMAC beads twice using 200 µL IMAC Loading Buffer (*see* **Note 5**).
3. Resuspend the beads in 200 µL IMAC Loading Buffer and add the sample (*see* **Note 6**).
4. Incubate the sample with IMAC beads in a Thermomixer for 30 min at room temperature.

5. Generate an IMAC micro-column essentially as described in Chapter 10.
6. Squeeze the tip of a 200 µL GELoader tip to prevent the IMAC beads from leaking.
7. After incubation, pack the beads in the constricted end of the GELoader tip by application of air pressure forming an IMAC micro-column [19].
8. It is critical to collect the IMAC flow-through (FT) in a new 1.7 mL low-binding microcentrifuge tube for further enrichment by TiO_2 chromatography (*see* Subheading 3.3).
9. Wash the IMAC column using 70 µL IMAC Loading Buffer.
10. Elute the mono-phosphorylated peptides bound to the IMAC beads using 80 µL of IMAC Elution Buffer 1. Collect the eluate into the IMAC-FT tube. The IMAC-FT and 1 % TFA elution fractions can be analyzed separately.
11. Pool the eluate with the IMAC-FT to obtain the SIMAC-mono fraction and lyophilize it prior to TiO_2 enrichment (*see* Subheading 3.3).
12. Elute the multi-phosphorylated peptides bound to the IMAC micro-column using 80 µL of IMAC Elution Buffer 2 directly into a p200 pipette tip containing a Poros Oligo R3 microcolumn (approximately 1 cm long).
13. Acidify with 100 % formic acid, typically 1 µL per 10 µL eluate (pH should be ~2–3), and 5 µL 100 % TFA, and desalt/concentrate the eluted multi-phosphorylated peptides on the Poros Oligo R3 micro-column (*see* Subheading 3.4).
14. Elute the peptides from the column using 60 µL RP Elution Buffer into a fresh 1.7 mL low binding microcentrifuge tube.
15. Lyophilize the sample prior to LC-MS/MS.

3.3 TiO_2 Batch Mode Purification of the "Mono"-Phosphorylated Peptides

1. Add acetonitrile, TFA, and glycolic acid to the SIMAC-mono peptide fraction to obtain TiO_2 Loading Buffer conditions (80 % acetonitrile, 5 % TFA, and 1 M glycolic acid) (*see* **Note 7**) or dilute the sample at least 10× with the TiO_2 Loading Buffer.
2. Add 0.6 mg TiO_2 beads per 100 µg peptide solution (*see* **Note 8**).
3. Place the tubes on a shaker (highest shaking) at room temperature for 5–10 min.
4. After incubation, centrifuge to pellet the beads (table centrifuge <15 s).
5. Transfer the supernatant to another low-binding tube and incubate it with another round of TiO_2 beads using half of the amount of TiO_2 beads as used in the first incubation. This can be repeated to recover larger amounts of phosphopeptides.

6. Pool the TiO_2 beads from the incubations using 100 µL Loading Buffer and transfer the solution to a new low-binding microcentrifuge tube (*see* **Note 9**).
7. Vortex the solution for 10 s and then centrifuge in a table centrifuge to pellet the beads. Remove the supernatant.
8. Wash the beads with 70–100 µL (*see* **Note 10**) Washing Buffer 1, mix for 10 s and then centrifuge to pellet the beads.
9. Wash the beads with 70–100 µL Washing Buffer 2, mix for 10 s and then centrifuge to pellet the beads. This step is important to remove peptides that bind to TiO_2 in a HILIC mode (*see* **Note 11**).
10. Dry the beads for 5–10 min in the vacuum centrifuge or on the table.
11. Elute the phosphopeptides with 100–200 µL Elution Buffer—mix well and leave the solution on a shaker for 10 min to allow an efficient elution.
12. Centrifuge the solution for 1 min and pass the supernatant over a small stage tip filter [20] (C8 stage tip) into a new low-binding tube to recover the liquid without any TiO_2 beads.
13. Wash the beads with 30 µL Elution Buffer and pool the wash (eluate) with the eluate from the previous step.
14. Elute potential bound peptides from the C8 filter with 5 µL 30 % acetonitrile and pool with the eluate from **steps 12** to **13**.
15. Lyophilize the eluted peptides or acidify the eluate with 1 µL formic acid per 10 µL eluate for direct cleanup of the phosphopeptides using RP material prior to downstream analyses as described for the multi-phosphorylated peptides above (*see* Subheading 3.4) (e.g., HILIC fractionation [17]).

3.4 Poros Oligo R3 Reversed Phase (RP) Micro-column Desalting/ Concentration of the Sample

Use GELoader tip micro-columns of ~6–10 mm or p200 pipette tips micro-columns (1–2 cm) depending on the amount of material to be purified. Here, it is illustrated for the p200 pipette tip (150 µg peptides from HeLa cell lysate).

1. Suspend Poros Oligo R3 reversed phase (RP) material in 200 µL 100 % acetonitrile.
2. Prepare a p200 pipette tip micro-column by stamping out a small plug of C_{18} material from a 3 M Empore™ C_{18} extraction disk and place it in the constricted end of the tip.
3. Pack Poros Oligo R3 RP beads on top of the p200 stage tip until the size of the column is 1–2 cm.
4. Load the acidified phosphopeptide sample slowly onto the RP micro-column (~1 drop/s).

5. Wash the RP micro-column using 60 µL RP Washing Buffer.
6. Elute the phosphopeptides from the RP micro-column using 40–60 µL RP Elution Buffer, followed by lyophilization of the phosphopeptides. (N.B. For MALDI MS analysis the peptides can be eluted off the GeLoader tip RP micro-column directly onto the MALDI target using 1 µL DHB solution. After crystallization the sample is ready for MALDI MS analysis).
7. Redissolve the lyophilized phosphopeptides in 0.5 µL 100 % formic acid and dilute immediately to 10 µL with UHQ water. The sample is then ready for LC-ESI-MS^n analysis.

3.5 µHPLC Tandem Mass Spectrometry (LC-MS/MS) Analysis

For LC-MS/MS analysis of purified phosphopeptides a standard strategy as described below can be used. A typical nanoLC setup would include a 0.075 mm × 200 mm analytical column packed with 3 µm RP resin interfaced with a high resolution/mass accuracy mass spectrometer as described in our original paper [17]. The number of phosphopeptides identified in the analysis can be increased by maximizing the resolution of the nanoLC separation via longer columns (e.g., 50 cm) and smaller chromatographic particle sizes (e.g., 1.9 µm). Alternatively, a two column system can be utilized using a 0.1 mm × 20 mm pre-column packed with RP resin (3–5 µm) combined with an analytical column as described above. A two column system is described below.

1. The phosphopeptides are redissolved in 0.1 % TFA and loaded onto a pre-column as described above using a µHPLC system (e.g., Dionex or EASY-LC) at a loading speed of 5 µL/min.
2. The phosphopeptides are eluted directly onto the analytical column (e.g., 0.075 mm × 200 mm) using a gradient (60–120 min) from 0 to 35 % B-Buffer (e.g., A-Buffer: 0.1 % formic acid; B-Buffer: 90 % acetonitrile, 0.1 % TFA) at an elution speed of 2–300 nL/min.
3. The phosphopeptides are eluted directly into a tandem mass spectrometer and analyzed by Data Dependent Analysis.

LC-ESI-MS/MS analysis of multi-phosphorylated peptides is improved by redissolving the phosphopeptides by sonication in an EDTA containing buffer prior to LC-ESI-MS/MS analysis [21].

An example of the results obtained by the SIMAC method using a relatively low complexity sample consisting of tryptic peptides derived from 12 standard proteins is shown in Fig. 2. The Figure shows the MALDI MS results obtained on a Bruker Ultraflex from a direct analysis of 1 pmol of the tryptic digest (Fig. 2a), the MALDI MS peptide mass map from the purification of the IMAC flow-through from 1 pmol peptide mixture using TiO_2 chromatography (Fig. 2b), the MALDI MS peptide mass map of the monophosphorylated peptides eluted from the IMAC material using 1 %

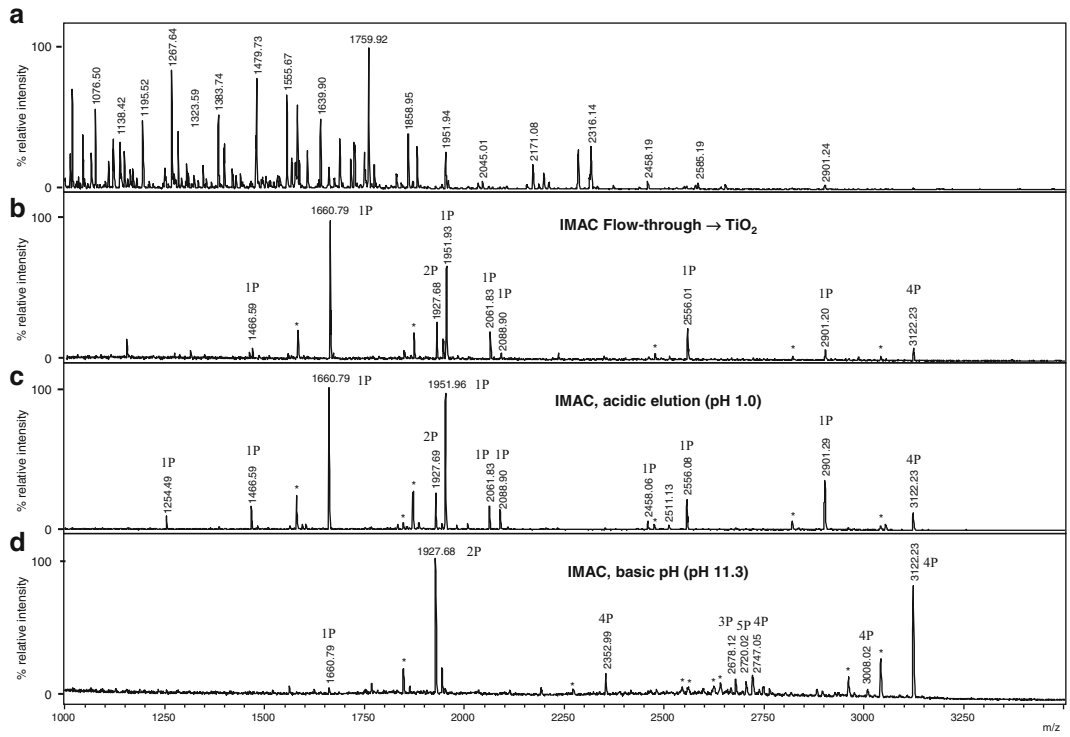


Fig. 2 Results obtained from 1 pmol peptide mixture using the SIMAC strategy. (**a**) MALDI MS peptide mass map of the direct analysis of the tryptic peptides. (**b**) MALDI MS peptide mass map of peptides identified from the IMAC flow-through after further enrichment using TiO_2 chromatography. (**c**) MALDI MS peptide mass map of peptides eluted from the IMAC micro-column using 1 % TFA. (**d**) MALDI MS peptide mass map of peptides eluted from the IMAC microcolumn using ammonia water (pH 11.30). The number of phosphate groups on the individual phosphopeptides is indicated by "#P". *Asterisk* indicates the metastable loss of phosphoric acid

TFA (Fig. 2c) and the MALDI MS peptide mass map obtained from the basic elution from the IMAC material (Fig. 2d). The phosphopeptides are illustrated by #P (*see* **Note 12**).

An example of the results obtained using the present SIMAC protocol for enrichment of phosphopeptides from a total of 150 µg peptides derived by tryptic digestion from a HeLa cell lysate is shown in Fig. 3. The enriched phosphopeptides were separated on a Dionex 3000 ultimate LC system using a homemade RP capillary column (25 cm) directly into a Q-Exactive Plus ESI-MS/MS instrument. The peptides were separated using a 90 min gradient from 0 to 25 % B Buffer (90 % acetonitrile in 0.1 % formic acid). The MS instrument was set to isolate and fragment 12 parent ions per MS cycle (MS and MS/MS resolution was set to 70,000 and 35,000 at 200 *m/z*, respectively; MS and MS/MS AGC target was 1E6 and 5E4, respectively; normalized collision energy was 30; isolation window was 1.5 Da). Here a total of 3370 unique phosphopeptides were identified from the 150 µg of starting material, using the Proteome Discoverer 1.4.1.14 (SwissProt_2014_04

(20340 entries)) with an enrichment percentage of about 88 % phosphopeptides using TiO_2 only (*see* Fig. 3a). When SIMAC was applied to the same sample a total of 5337 unique phosphopeptides (enrichment percentage 89 %) could be identified, whereof 3804 and 2499 were identified in the SIMAC mono and multi fractions, respectively. Of these, only 966 unique phosphopeptides were shared between the two fractions (*see* Fig. 3b) indicating a good separation. When looking at the number of phosphate groups on the unique phosphopeptides identified in each fraction a clear enrichment of multi-phosphorylated peptides could be seen when using the SIMAC procedure, as the SIMAC multi fraction contained 56.5 % phosphopeptides (*see* Fig. 3d) with 2 or more phosphate groups compared to only 14 % in the mono fraction (Fig. 3c) and 22 % in the TiO_2 enrichment (*see* Chapter 9). In total the SIMAC procedure resulted in the identification of 31 % multi-phosphorylated peptides.

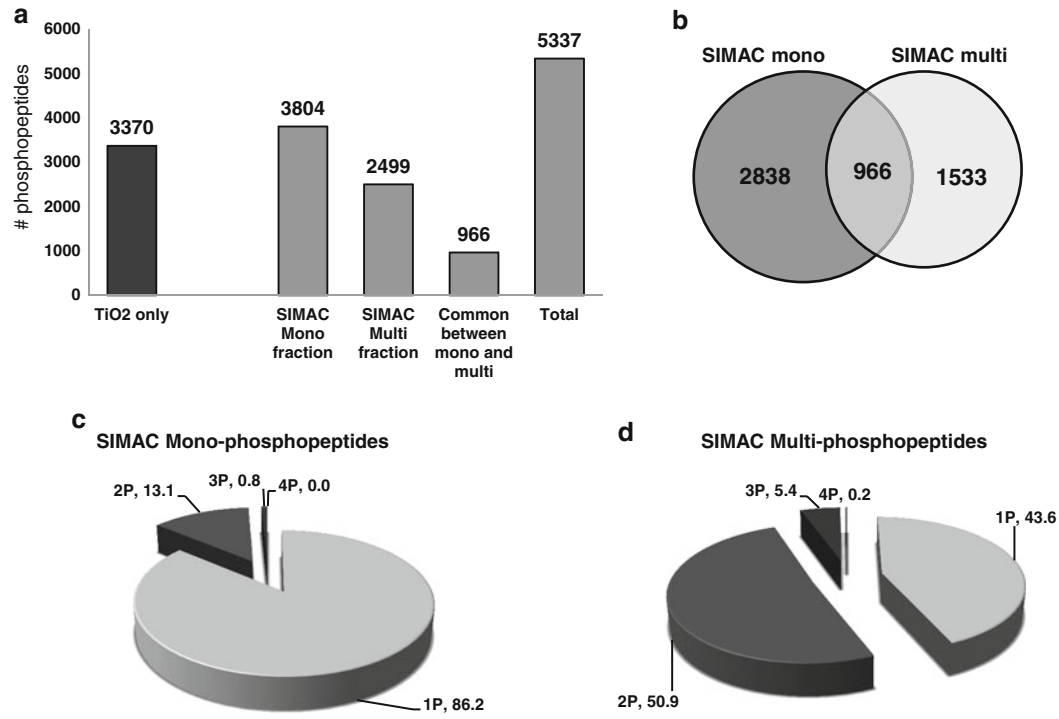


Fig. 3 Results obtained from the enrichment of phosphorylated peptides from acetone precipitated proteins from HeLa cells using TiO_2 chromatography or SIMAC. (**a**) Overview of the number of unique phosphopeptides identified in the TiO_2 and SIMAC experiments. (**b**) Venn diagram showing the overlap between the SIMAC mono and multi fractions. (**c**) Percentage distribution of the number of phosphate groups on the phosphopeptides identified in the SIMAC mono fraction. (**d**) Percentage distribution of the number of phosphate groups on the phosphopeptides identified in the SIMAC multi fraction

4 Notes

1. It is important to obtain the highest purity of all chemicals used.
2. All solutions should be prepared in UHQ water.
3. Always start by testing the method using a model peptide mixture. It is important to freshly prepare the peptide mixture as peptides bind to the surface of the plastic tubes in which they are stored. In addition, avoid transferring the peptide sample to different tubes to minimize adsorptive losses of the sample.
4. The peptide mixture used for the experiment illustrated in this chapter contained peptides originating from tryptic digestions of 1 pmol of each of the 12 proteins. Experiments have shown that the presented method is sensitive down to the low femtomole level [11].
5. The PhosSelect IMAC beads are very fragile so high speed mixing should be avoided in any steps.
6. The sample should be diluted in IMAC Loading Buffer or for larger volume add 100 % TFA and 100 % acetonitrile to make the sample up to the IMAC Loading Buffer. The total volume should not exceed 300 µL.
7. If you have 100 µL peptide sample, you can add 50 µL water, 50 µL 100 % TFA, 800 µL acetonitrile, and 76 mg glycolic acid to make the sample up to the proper TiO_2 Loading Buffer.
8. The optimal amount of TiO_2 beads to add to the sample in order to reduce non-specific binding and optimize phosphopeptide yield is 0.6 mg TiO_2 per 100 µg of peptide starting material (*see* [17] for further information). This will of course change depending on the source of biological material used as TiO_2 selectively enriches other biomolecules (reviewed in [22]) such as sialylated glycopeptides [23] and acidic lipids [24] commonly found in membrane fractions.
9. The transfer to a new tube is performed due to the fact that peptides stick to plastic and can be eluted from the plastic surface in the last elution step resulting in contamination with non-modified peptides.
10. For larger scale analysis, where more TiO_2 beads are used, larger volumes of the buffers should be used.
11. TiO_2 is an efficient HILIC material and hydrophilic peptides can bind to the material when loaded in high organic solvent. The inclusion of 5 % TFA and 1 M glycolic acid should prevent most hydrophilic non-modified peptides from binding, however, some can still be found in the eluates from TiO_2.

Therefore in order to eliminate any binding from non-modified hydrophilic peptides this last Washing Buffer is important. For membrane preparations the last washing supernatant will contain neutral glycopeptides which can then be analyzed further.

12. The results obtained using this protocol will differ according to the mass spectrometer used for the analysis of the phosphopeptides, not only between MALDI MS and ESI MS but also within different MALDI MS instruments, depending on laser optics, laser frequency, instrumental Configuration, sensitivity, etc.

Acknowledgements

This work was supported by the Danish Natural Science and Medical Research Councils (grant no. 10-082195 (T.E.T)) and the Lundbeck Foundation (M.R.L—Junior Group Leader Fellowship).

References

1. Li S, Dass C (1999) Iron(III)-immobilized metal ion affinity chromatography and mass spectrometry for the purification and characterization of synthetic phosphopeptides. Anal Biochem 270:9–14
2. Neville DC, Rozanas CR, Price EM, Gruis DB, Verkman AS, Townsend RR (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. Protein Sci 6:2436–2445
3. Posewitz MC, Tempst P (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. Anal Chem 71:2883–2892
4. Kuroda I, Shintani Y, Motokawa M, Abe S, Furuno M (2004) Phosphopeptide-selective column-switching RP-HPLC with a titania precolumn. Anal Sci 20:1313–1319
5. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4:873–886
6. Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76:3935–3943
7. Sano A, Nakamura H (2004) Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. Anal Sci 20:565
8. Zhou HL, Watts JD, Aebersold RA (2001) A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol 19:375–378
9. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods 4:231–237
10. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101:12130–12135
11. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC - a phosphoproteomic strategy for the rapid separation of mono-phosphorylated from multiply phosphorylated peptides. Mol Cell Proteom 7(4):661–671
12. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. Anal Chem 76: 3590–3598
13. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods 4:709–712
14. Chalmers MJ, Hakansson K, Johnson R, Smith R, Shen J, Emmett MR, Marshall AG (2004) Protein kinase A phosphorylation characterized by tandem Fourier transform ion cyclotron

resonance mass spectrometry. Proteomics 4:970–981

15. Schroeder MJ, Webb DJ, Shabanowitz J, Horwitz AF, Hunt DF (2005) Methods for the detection of paxillin post-translational modifications and interacting proteins by mass spectrometry. J Proteome Res 4:1832–1841
16. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol 20:301–305
17. Engholm-Keller K, Birck P, Storling J, Pociot F, Mandrup-Poulsen T, Larsen MR (2012) TiSH--a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. J Proteomics 75:5749–5761
18. Engholm-Keller K, Hansen TA, Palmisano G, Larsen MR (2011) Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO(2) chromatography applied to proximal EGF signaling. J Proteome Res 10:5383–5397
19. Gobom J, Nordhoff E, Mirgorodskaya E, Ekman R, Roepstorff P (1999) Sample purification and preparation technique based on nanoscale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom 34:105–116
20. Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75:663–670
21. Liu S, Zhang C, Campbell JL, Zhang H, Yeung KK, Han VK, Lajoie GA (2005) Formation of phosphopeptide-metal ion complexes in liquid chromatography/electrospray mass spectrometry and their influence on phosphopeptide detection. Rapid Commun Mass Spectrom 19:2747–2756
22. Engholm-Keller K, Larsen MR (2011) Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds--applications in acidic modification-specific proteomics. J Proteomics 75:317–328
23. Jensen SS, Larsen MR (2007) Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. Rapid Commun Mass Spectrom 21:3635–3645
24. Calvano CD, Jensen ON, Zambonin CG (2009) Selective extraction of phospholipids from dairy products by micro-solid phase extraction based on titanium dioxide microcolumns followed by MALDI-TOF-MS analysis. Anal Bioanal Chem 394:1453–1461
25. Thingholm TE, Jensen ON, Larsen MR (2009) Enrichment and separation of mono- and multiply phosphorylated peptides using sequential elution from IMAC prior to mass spectrometric analysis. Methods Mol Biol 527:67–78, xi

Chapter 11

Improving the Phosphoproteome Coverage for Limited Sample Amounts Using TiO_2-SIMAC-HILIC (TiSH) Phosphopeptide Enrichment and Fractionation

Kasper Engholm-Keller and Martin R. Larsen

Abstract

Obtaining high phosphoproteome coverage requires specific enrichment of phosphorylated peptides from the often extremely complex peptide mixtures generated by proteolytic digestion of biological samples, as well as extensive chromatographic fractionation prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Due to the sample loss resulting from fractionation, this procedure is mainly performed when large quantities of sample are available. To make large-scale phosphoproteomics applicable to smaller amounts of protein we have recently combined highly specific TiO_2-based phosphopeptide enrichment with sequential elution from immobilized metal affinity chromatography (SIMAC) for fractionation of mono- and multi-phosphorylated peptides prior to capillary scale hydrophilic interaction liquid chromatography (HILIC) based fractionation of monophosphorylated peptides. In the following protocol we describe the procedure step by step to allow for comprehensive coverage of the phosphoproteome utilizing only a few hundred micrograms of protein.

Key words Phosphopeptide, Phosphoproteomics, Protein phosphorylation, TiO_2, Capillary HPLC, HILIC, TiSH, Enrichment, Fractionation

1 Introduction

Mass spectrometry (MS)-based proteomics have improved tremendously within the last decade, leading to a rapid progression in our knowledge of cellular signaling processes. Particularly in the field of post-translational modifications (PTMs), MS has become an indispensable tool for large-scale characterization and quantification of protein phosphorylation, glycosylation, ubiquitination, acetylation, etc. As a result of the generally low stoichiometry of these PTMs, specific enrichment steps are absolutely necessary to achieve adequate coverage of the modified sites in the samples of interest. While immobilized metal affinity chromatography (IMAC) initially was the method of choice for phosphopeptide enrichment [1, 2], TiO_2-based enrichment strategies [3–7] have in many laboratories

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_11,

become the de facto standard in phosphoproteomics due to their high specificity and robustness towards detergents and salts [8]. Furthermore, the combination of the two enrichment techniques (in a procedure called "*S*equential elution from *IMAC*"—SIMAC) has been shown to facilitate separation of mono- and multi-phosphorylated peptides [9]. However, in spite of increasingly sensitive and fast MS instruments facilitating identification of thousands of phosphorylation sites in a few hours, sample complexity still poses a challenge in the quest for achieving a complete coverage of the cellular phosphoproteome. Five years ago hydrophilic interaction liquid chromatography (HILIC) was introduced as a prefractionation step in large-scale phosphoproteomics [10] and is now, along with strong cation exchange and high pH reversed phase chromatography, commonly used as the first dimension of fractionation prior to reversed phase (RP) nanoLC-MS/MS analysis of phosphopeptides. Amongst the advantages of HILIC are a salt-free buffer system and high orthogonality with the second dimension RP LC separation [11]. However, HILIC has been shown to deliver a lower resolution for multi-phosphorylated peptides than for singly phosphorylated species [12, 13]. Furthermore, as the first dimension of chromatography traditionally has been performed prior to phosphopeptide enrichment, the process of enriching all the collected fractions lowers the throughput of the setup. Moreover, fractionating the small amounts of purified phosphopeptides using standard HPLC columns (2–4.6 mm i.d.) can lead to higher phosphopeptide losses due to adsorption to column and plastic surfaces. Performing peptide fractionation prior to phosphopeptide enrichment can also complicate the experimental workflow if the highest possible specificity is aspired: The ratio of TiO_2 beads to peptide starting material is very important for obtaining a selective purification of phosphopeptides and the addition of too much TiO_2 resin to a peptide fraction of unknown quantity will lead to a significant co-purification of non-modified peptides [12, 14].

To circumvent these drawbacks, while still utilizing the advantages of TiO_2, SIMAC, and HILIC, we have combined all three enrichment/fractionation methods into an integrated method dubbed TiSH—*Ti*O_2-*S*IMAC-*H*ILIC [12] (*see* Fig. 1). The TISH-strategy relies on a TiO_2-based pre-enrichment step, removing most non-phosphorylated peptides prior to separation of mono- and multi-phosphorylated peptides. The sample containing multi-phosphorylated peptides is less complex and contains a smaller amount of peptide and is therefore analyzed directly by RP nanoLC-MS/MS. The sample containing mono-phosphorylated peptides is fractionated using capillary scale HILIC (320 μm i.d. in-house made column (*see* Fig. 2)). Fraction collection is performed directly into a 96-well microplate, in which the samples can be dried by vacuum centrifugation and dissolved in an RP compatible buffer and directly analyzed by nanoLC-MS/MS. Using

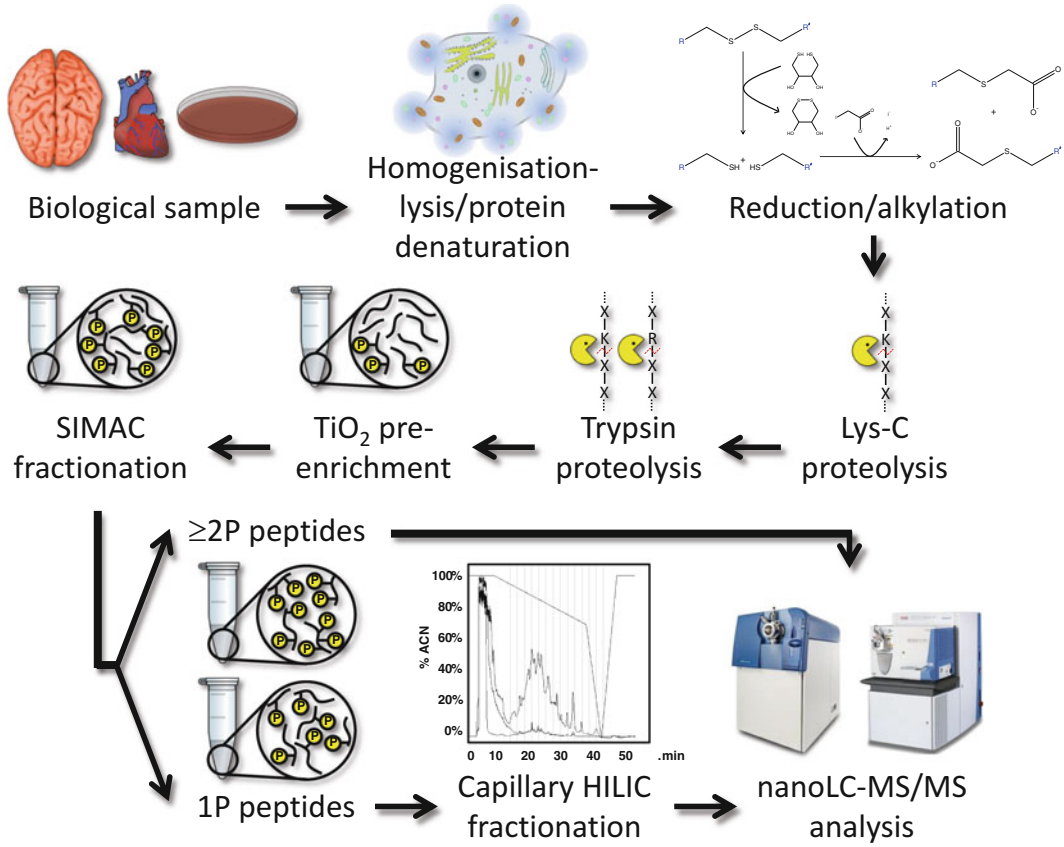


Fig. 1 Outline of the overall TiSH phosphopeptide enrichment and fractionation strategy. After an initial integrated homogenization/lysis-protein denaturation of the tissue or cells, the protein sample is reduced and alkylated prior to proteolytic digestion using endoproteinase Lys-C, followed by tryptic proteolysis. The resulting peptides are subjected to an initial TiO_2 enrichment to remove most non-phosphorylated peptides before the SIMAC fractionation of mono-phosphorylated and multi-phosphorylated peptides. While the latter are analyzed directly by LC-MS/MS, the mono-phosphorylated peptides are further fractionated using capillary HILIC before LC-MS/MS analysis

this setup, 6600 unique phosphopeptides could be identified in a quantitative duplex dimethylation experiment with starting amounts of 300 µg of insulinoma cell line protein per condition, which was analyzed on an LTQ-Orbitrap Velos using a standard (0.075 mm × 20 mm column) nanoLC setup [12].

The exact procedure for cell lysis and protein extraction is highly dependent on the nature and amount of biological sample. The lysis and digestion procedure presented here is suited for low to medium amounts (100 µg to 2 mg protein) of soft tissue or cultured cells and is focused on minimizing sample losses by avoiding protein precipitation steps etc., but can be substituted with other lysis/extraction protocols if necessary as TiO_2-based phosphopeptide enrichment is highly robust towards various salts, buffers and denaturing reagents [8].

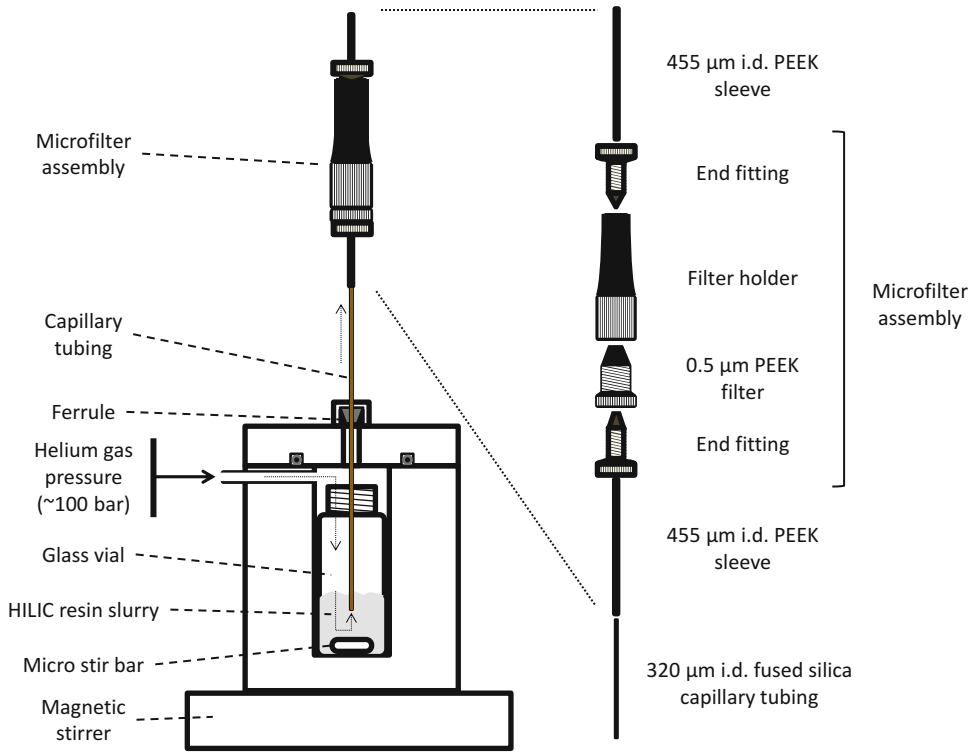


Fig. 2 Schematic illustration of the in-house made capillary HILIC column and the column packing process. The column consists of a 0.5 µm PEEK micro filter connected to a 320 µm i.d. fused silica capillary using a 455 µm i.d. sleeve. The column is packed with the TSKGel Amide-80 HILIC resin slurry from a vial in a high-pressure packing device into which the upstream end of the fused silica capillary is inserted. A magnetic stirrer underneath the device and a micro stir bar in the vial keeps the resin in suspension and by applying a helium gas pressure (~100 bar) to the device, forcing the slurry into the capillary, the resin is trapped by the micro filter and the column packed. When the column has packed to the desired length, the pressure can be released very slowly, and the column removed and attached to the capillary HPLC system

When processing protein amounts outside the 0.1–2 mg range, adjusting the quantities of chromatographic resin/columns, volumes, etc. is necessary for optimal efficiency and specificity of the protocol. For a simpler and faster procedure, the protocol can be performed without the SIMAC mono-/multi-phosphopeptide fractionation step, but will potentially result in a lower number of multi-phosphorylated peptide identifications.

2 Materials

All solutions are prepared using ultrapure water and analytical grade chemical reagents. To minimize protein and peptide adsorption to plastics surfaces, all microcentrifuge tubes and tips should be of the "low-binding" type.

2.1 Cell Lysis, Protein Extraction and Digestion

1. Ultrapure water (18 MΩ-cm).
2. 8 M urea.
3. Phosphatase Inhibitor Solution.
4. 1 M dithiothreitol (DTT).
5. 0.5 M iodoacetamide (IAA).
6. Ultrasound probe sonicator with micro probe (e.g., Branson 250A Sonifier Ultrasonic Processor Cell Disruptor with Micro Tip).
7. 1 M triethyl ammonium bicarbonate (TEAB) (*see* **Note 1**).
8. Protein concentration measurement method such as amino acid composition analysis or other technique capable of measuring low microgram amounts of protein/peptides such as the Pierce Micro BCA Protein Assay Kit (Thermo Fisher Scientific) (*see* **Note 2**).
9. 0.05 AU/μL endoproteinase Lys-C in H_2O.
10. 1 μg/μL Sequencing Grade Modified Trypsin in 50 mM TEAB (*see* **Note 3**).
11. 10 % trifluoroacetic acid (TFA).

2.2 Phosphopeptide Pre-enrichment

1. Ultrapure water (18 MΩ-cm).
2. Acetonitrile (ACN, HPLC grade).
3. Trifluoroacetic acid (TFA).
4. Methanol (MeOH).
5. Titansphere 5 μm TiO_2 resin (GL Sciences).
6. Glycolic acid.
7. Ammonium hydroxide 25 % (NH_4OH).
8. TiO_2 Loading Buffer: 1 M glycolic acid in 80 % ACN, 5 % TFA.
9. TiO_2 Washing Buffer 1: 80 % ACN, 1 % TFA.
10. TiO_2 Washing Buffer 2: 20 % ACN, 0.2 % TFA.
11. TiO_2 Elution Buffer: 1 % NH_4OH, pH 11.

2.3 Separation of Multi- and Mono-phosphorylated Peptides Using Sequential Elution from IMAC (SIMAC)

1. PhosphoSelect IMAC resin (Sigma-Aldrich).
2. 200 μL gel loader tips.
3. 1 mL plastic syringe with home-made adaptor facilitating attachment to gel loader tip.
4. Formic acid (FA).
5. Titansphere 5 μm TiO_2 resin (reuse beads from phosphopeptide pre-enrichment) (*see* Subheading 3.2, **step 9**).
6. Trifluoroacetic acid (TFA).
7. Acetonitrile (ACN, HPLC grade).

8. SIMAC Loading Buffer: 50 % ACN, 0.1 % TFA.
9. SIMAC Acid Elution Buffer: 20 % ACN, 1 % TFA (*see* **Note 4**).
10. SIMAC Basic Elution Buffer: 1 % NH_4OH, pH 11.
11. pH meter with micro pH electrode.
12. Gauge 25 blunt HPLC syringe needle.
13. POROS Oligo R3 (Life Technologies).
14. 3 M Empore C18 extraction disc (3 M).
15. Reversed Phase (RP) Conditioning/Elution Buffer: 70 % ACN, 0.1 % TFA.
16. Reversed Phase (RP) Washing Buffer: 0.1 % TFA.
17. Vacuum centrifuge.

2.4 Desalting of Non-phosphorylated Peptides

1. Sep-Pak C18 Plus Light cartridge (Waters).
2. Reversed Phase (RP) Conditioning/Elution Buffer: 70 % ACN, 0.1 % TFA.
3. Reversed Phase (RP) Washing Buffer: 0.1 % TFA.

2.5 Packing of Hydrophilic Interaction Liquid Chromatography (HILIC) Capillary Column

1. TSKGel Amide 80 3 μm HILIC resin (Tosoh Bioscience, from HPLC column).
2. Methanol (MeOH).
3. Polyether ether ketone (PEEK) inline microfilter (Upchurch Scientific).
4. MicroTight Sleeve Black (F-186), 455 μm i.d. (Upchurch Scientific).
5. 0.32 mm (inner diameter (i.d.))/0.435 mm (outer diameter (o.d.)) fused silica capillary tubing (Polymicro Technologies).
6. Vertical high-pressure capillary column packing device (e.g. Nanobaume (Western Fluidics) or PC77 Pressure Injection Cell (Next Advance, Inc.)).
7. Vespel ferrule (0.5 mm (i.d.) × 3.2 mm (o.d.) × 3.7 mm) (SGE Analytical Science).
8. 2 mL glass vial.
9. Micro stir bar.
10. Magnetic stirrer.
11. Compressed helium.

2.6 Peptide Fractionation via Hydrophilic Interaction Liquid Chromatography (HILIC)

1. Capillary flow (1–20 μL/min) high-performance liquid chromatography system with UV detector. In this protocol an Agilent 1200 (Agilent Technologies) equipped with an autosampler with a 40 μL loop, a UV detector with an 80 nL flow cell and a micro fraction collector was used.
2. In-house made 0.32 × 20 mm 3 μm resin TSKGel Amide 80 HILIC capillary column.

3. Dimethyl sulfoxide (DMSO).
4. Ultrapure water (18 MΩ-cm).
5. Acetonitrile (ACN, HPLC grade).
6. Trifluoroacetic acid (TFA).
7. HILIC Solvent A: 0.1 % TFA.
8. HILIC Solvent B: 90 % ACN, 0.1 % TFA.

2.7 Analysis by Mass Spectrometry

1. Mass spectrometer capable of performing MS/MS—preferentially a high-resolution/high mass accuracy instrument (recent Q-TOF, or Orbitrap based mass spectrometer) interfaced to a nanoHPLC with a 50–100 μm i.d. RP capillary column setup for highly sensitive online peptide separation.
2. Software for processing of raw mass spectrometry data files and generation of peak lists for searching against a protein database (e.g. Uniprot) such as Mascot/Mascot Distiller (Matrix Science, London, UK) (data from most vendors and instruments), Proteome Discoverer (Thermo Scientific, Bremen, Germany) (data from Thermo instruments), MaxQuant [15] (high resolution data from Thermo Orbitrap instruments and certain Bruker and ABSciex Q-TOFs) and the TransProteomicPipeline [16] (vendor independent).

3 Methods

3.1 Cell Lysis, Protein Extraction and Digestion

1. Prepare lysis buffer by making an 8 M urea solution in 50 mM TEAB and add phosphatase inhibitors to a 1× final concentration. Add the minimum amount of lysis buffer necessary for full lysis of the cells on ice. Keeping the volume down is important as the sample has to be diluted to ≤1 M urea (≥8× dilution) prior to tryptic digestion. Vortex extensively until a clear solution is obtained.
2. If sample volume is large enough to allow for sonication: Sonicate the sample for 3 × 10 s (with 10 s breaks for the sample to cool) on ice (*see* **Note 5**) at the highest intensity possible without causing the sample to foam.
3. Measure protein concentration of a small aliquot of the sample while keeping the rest of the sample on ice to reduce endogenous protease activity.
4. Perform combined reduction of protein disulfide bonds and Lys-C proteolysis by adding 1 M DTT to a final concentration of 10 mM and Lys-C (0.05 AU Lys-C per 100–300 μg) to the sample and incubate at room temperature (*see* **Note 5**) for 3 h.
5. Alkylate reduced cysteine residues by adding 0.5 M IAA to a final concentration of 20 mM, incubating the sample in the dark at room temperature for 20 min.

6. Dilute the sample 8× in 50 mM TEAB to a final concentration of 1 M urea.
7. Optional—if no sonication was performed in **step 2**: Sonicate the sample on ice (*see* **Note 5**).
8. Add trypsin in a 1:50 trypsin–protein ratio and incubate the sample at room temperature (*see* **Note 5**) overnight.
9. At this step the sample(s) can be subjected to chemical stable isotope labeling such as reductive dimethylation [17], iTRAQ [18] or TMT [19] for relative quantification (*see* **Note 1**).
10. Adjust the sample to a pH of below 3 using 10 % TFA or FA, centrifuge the sample at 14,000×*g* to precipitate lipids and transfer the supernatant to a new 1.5 mL microcentrifuge tube.
11. Optionally, at this step a fraction of the sample can be removed for later LC-MS/MS analysis if protein-level identification/quantification is required. Alternatively, non-phosphorylated peptides can be obtained from the TiO_2 pre-enrichment flow-through fraction (*see* Subheading 3.4 and **Note 6**).

3.2 TiO_2 Pre-enrichment

1. Adjust the peptide sample to 1 M glycolic acid in 80 % ACN, 5 % TFA. For example, if the sample is present in 150 µL add 50 µL 100 % TFA, 800 µL ACN and 76.05 mg of glycolic acid (*see* **Note 7**). Vortex to dissolve the glycolic acid.
2. Add TiO_2 resin—0.6 mg per 100 µg of peptide—and incubate under vigorous shaking for 10 min. Pellet the resin by brief centrifugation (2000×*g* for 1 min) and transfer the supernatant to a new tube (Optional: If enriching for other PTMs after phosphopeptide enrichment or analysis of the non-modified peptides, save the flow-through for later desalting in Subheading 3.4).
3. Optional second incubation for improved recovery of phosphopeptides: Add TiO_2 resin—0.3 mg per 100 µg of peptide—to the supernatant in the new tube and incubate under vigorous shaking for 10 min. Pellet the resin by brief centrifugation (2000×*g* for 1 min) and transfer the supernatant containing non-phosphorylated peptides to a new tube (can be stored at −80 °C or directly subjected to vacuum centrifugation prior to desalting)—(*see* **Note 6** and Subheading 3.4).
4. Add 500 µL of TiO_2 Loading Buffer to the first tube containing the TiO_2 resin and mix by vortexing. If a second incubation was performed (**step 3**), transfer the suspension to the tube containing the TiO_2 resin and transfer the resin pool to a new 1.5 mL micro centrifuge tube to leave behind non-phosphorylated peptides absorbed to the tube.
5. Vortex briefly and pellet the resin by brief centrifugation (2000×*g* for 1 min). Discard the supernatant.

6. Wash the resin in 500 µL TiO_2 Washing Buffer 1, vortex briefly, and pellet the resin by brief centrifugation (2000 × *g* for 1 min). Discard the supernatant.
7. Wash the resin in 500 µL TiO_2 Washing Buffer 2, vortex briefly, and pellet the resin by brief centrifugation (2000 × *g* for 1 min). Discard the supernatant, and dry the resin briefly in a vacuum centrifuge to evaporate residual acidic buffer.
8. Elute the phosphorylated peptides off the TiO_2 resin by adding 100 µL TiO_2 Elution Buffer to the resin, and incubate the slurry with mixing for 15 min followed by 1 min of centrifugation at 14,000 × *g* (*see* **Note 8**). Transfer as much supernatant as possible to a new tube without aspirating any beads.
9. Add another 40 µL of TiO_2 Elution Buffer and 10 µL ACN to the resin, mix briefly and pellet the resin by centrifugation for 1 min at 14,000 × *g*. Aspirate as much of the supernatant as possible without disturbing the resin and pool the liquid with the first eluate (*see* **Note 9**). Wash the TiO_2 resin in 500 µL TiO_2 Washing Buffer 1, vortex briefly, and pellet the resin by brief centrifugation (2000 × *g* for 1 min). Discard the supernatant and save the resin for the later TiO_2 enrichment step (*see* **Note 10**).
10. Dry the eluate to completeness in a vacuum centrifuge (*see* **Note 11**).

3.3 Separation of Multi- and Mono-phosphorylated Peptides Using Sequential Elution from IMAC (SIMAC)

1. Redissolve the phosphopeptide sample in first 0.5 µL 100 % FA, followed by 200 µL 50 % ACN, 0.1 % TFA and adjust it to pH 1.8 with 10 % TFA, ideally measuring the pH using a micro electrode pH meter.
2. Wash 60 µL of IMAC slurry by adding 200 µL SIMAC Washing Buffer. Pellet the resin by centrifugation at 2000 × *g* for 15 s, discard the liquid and repeat the washing step. After having removed the liquid from the resin, add the phosphopeptide sample to the resin.
3. Incubate the sample under continuous shaking for 30 min at room temperature to allow the phosphopeptides to bind the IMAC resin. Prepare 200 µL gel loader tips by constricting the end of the tip to make it retain the IMAC resin. After incubation, centrifuge briefly and transfer most of the supernatant to a new 1.5 mL tube without disturbing the resin pellet and resuspend the beads in the leftover sample. Transfer the slurry to the gel loader tip. Push the liquid through the tip by applying air pressure using a plastic syringe and collect the flow-through, containing some mono-phosphorylated as well as non-phosphorylated peptides, in the same 1.5 mL tube containing the IMAC flow-through. Make sure that there are no beads in the flow-though (*see* **Note 12**).

4. Wash the IMAC column formed in the constricted end of the tip with 70 µL SIMAC Washing Buffer, washing off non-phosphorylated peptides as well as weakly binding mono-phosphorylated peptides and collect the wash in the tube containing the flow-through from **step 3**.
5. Elute the remaining mono-phosphorylated peptides slowly off the IMAC column (1 droplet/s (*see* **Note 13**)) by adding 70 µL SIMAC Acid Elution Buffer (*see* **Note 4**) to the column and applying air pressure using the syringe—pool this eluate with the flow-through/wash from **step 4** resulting in a combined mono-phosphorylated peptide sample. Make sure no IMAC resin is present in sample (*see* **Note 12**).
6. Slowly elute the multi-phosphorylated peptides off the IMAC resin into a microcentrifuge tube by adding 100 µL of SIMAC Basic Elution Buffer to the column and applying air pressure via the syringe. Make sure no IMAC beads are present in the eluate (*see* **Note 12**) and acidify the sample with 10 µL 100 % FA. Save the sample for later concentration/desalting prior to LC-MS/MS analysis.
7. Adjust the sample containing mono-phosphorylated peptides from **step 5** to 80 % ACN, 1 % TFA and incubate it under vigorous shaking for 10 min with the same amount of TiO_2 as used in the pre-enrichment procedure (*see* Subheading 3.2, **step 2**). The TiO_2 resin from Subheading 3.2, **step 9** can be reused (after regeneration) in this step (*see* **Note 10**).
8. After incubation, pellet the beads by centrifugation at 2000 × *g* for 1 min and transfer the supernatant to another tube (save the resin). Optional second enrichment (for improved recovery): The supernatant can be incubated with the same amount of TiO_2 resin as in Subheading 3.2, **step 3** for an additional 10 min under vigorous shaking. The resin used in Subheading 3.2, **step 3** can be reused (after regeneration) (*see* **Note 10**).
9. After the second incubation, pellet the resin by centrifugation at 2000 × *g* for 1 min, and discard the supernatant. Add 300 µL 50 % ACN, 0.1 % TFA to each of the resin-containing tubes, vortex briefly and combine the two slurries. Pellet the resin by centrifugation at 2000 × *g* for 1 min and discard the supernatant.
10. Dry the TiO_2 resin for 5 min in a vacuum centrifuge to remove leftover Washing Buffer and elute the mono-phosphorylated peptides by adding 100 µL TiO_2 Elution Buffer to the tube. Vortex briefly and incubate under continuous shaking for 15 min. Pellet the resin by centrifugation at 14,000 × *g* for 1 min and transfer the eluate to another tube without disturbing the resin bed. Add another 50 µL of TiO_2 Elution Buffer

to the resin, briefly vortex, centrifuge again at 14,000 × *g* for 1 min. Combine the second eluate with the first one without disturbing the resin bed. Make sure that no resin is present in the eluate as described in Subheading 3.2, **step 9**. Acidify the sample with 15 µL 100 % FA.

11. Prior to fractionation of the mono-phosphorylated peptides by HILIC and LC-MS/MS analysis of the multi-phosphorylated peptides both samples are desalted on homemade POROS R3 RP micro-columns (~1–2 cm long) packed in a 200 µL pipette tip constricted with a plug of 3 M Empore C18 disc.
12. Before applying the sample, wash the column in 100 µL RP Conditioning/Elution Buffer, followed by 100 µL RP Washing Buffer. Load the sample onto the column by applying air pressure, wash it with 100 µL RP Washing Buffer and elute off the phosphopeptides with 100 µL RP Conditioning/Elution Buffer into a new tube. Dry the phosphopeptides sample to completion in a vacuum centrifuge and store at −20 °C until time of fractionation/analysis.

3.4 Desalting of Non-phosphorylated Peptides

1. This step is optional depending on whether analysis of other post-translational modifications or the non-modified peptides is required; e.g. lysine acetylated peptides can be immunoprecipitated after desalting of the non-phosphorylated peptide sample. The Sep-Pak C18 Plus Light has a capacity of a least 1–2 mg peptide.
2. Dry the samples from Subheading 3.2, **step 2** to completeness in a vacuum centrifuge and dissolve the peptides in 3 mL 0.1 % TFA.
3. Wash the Sep-Pak cartridge with 3 mL RP Conditioning/Elution Buffer followed by washing with 5 mL of RP Washing Buffer.
4. Load the peptide solution onto the cartridge slowly (1 drop/s) and discard the flow-through. Wash the cartridge with 5 mL RP Washing Buffer. Flush the cartridge with air, removing the solvent completely.
5. Elute the peptides using 1 mL of RP Conditioning/Elution Buffer slowly (1 drop/s) into a new tube. Dry the sample by vacuum centrifugation and store at −20 °C until further use.

3.5 Packing of Hydrophilic Interaction Liquid Chromatography (HILIC) Capillary Column

1. Make a 0.5 mL 5–10 % slurry of HILIC resin in 100 % MeOH in a glass vial and sonicate it for 5 min in a sonicator bath, add the micro stir bar, and place it in the column packing device (*see* Fig. 2).
2. Cut a ~25 cm piece of 0.32 mm i.d. capillary tubing and attach the micro filter via the sleeve. Insert the other end of the capillary through the ferrule and attach it to the column packing

device (*see* Fig. 2). Connect the packing device to the compressed helium supply and adjust the pressure to 100 bars. Leave the column packing with magnetic stirring on overnight or until the column has packed. Leave to dry, as releasing the pressure from a wet column can lead to back-flow of the resin.

3. Release the pressure very slowly, remove the column from the packing device and connect it to the capillary HPLC system.

3.6 Peptide Fractionation via HILIC

The capacity of the HILIC column used in this protocol (0.32 × 200 mm) is at least 10–20 μg of peptide, corresponding to phosphopeptides originating from approximately 1–2 mg protein lysate. If larger amounts of sample are available, scale up the HILIC column and flow rate of the HPLC run.

1. Dissolve the mono-phosphorylated peptide sample in 0.2 μL 50 % DMSO (*see* **Note 14**). Add 3.6 μL H_2O, followed by 0.4 μL 10 % TFA, and finally slowly add 36 μL ACN (*see* **Note 14**). Centrifuge the sample at 14,000 × *g* for 3 min to precipitate any undissolved material.
2. Load the supernatant onto the HILIC column at a flow rate of 12 μL/min for 10 min, followed by separation of the phosphopeptides at a flow rate of 6 μL/min with an increasing aqueous gradient from 100 to 60 % B over 40 min. Collect 1 min fractions throughout the gradient into a 96-well microtiter plate.
3. Combine the fractions based on UV absorption at 214 nm to obtain a number of samples in the plate (typically 10–15—depending on time available for LC-MS/MS analysis) containing similar amounts of phosphopeptides (*see* **Note 15**). Dry the sample plate by vacuum centrifugation and dissolve the samples in 0.3 μL 100 % FA followed by 4.7 μL H_2O (for a 5 μL nanoLC-MS/MS injection).

3.7 Analysis by Mass Spectrometry

1. Analyze the samples by RP nanoLC-MS/MS. A typical nanoLC setup would include a 0.075 mm × 20 mm analytical column packed with 3 μm RP resin interfaced with a high resolution/mass accuracy mass spectrometer as described in our original paper [12]. The number of phosphopeptides identified in the analysis can be increased by maximizing the resolution of the nanoLC separation via longer columns (e.g. 50 cm) and smaller chromatographic particle sizes (e.g. 1.9 μm) using a nanoLC system capable of operating at 800–1000 bar. Depending on the speed of mass spectrometer, time available for LC-MS/MS analysis, complexity of the sample and length of nanoLC column, the samples should be analyzed using an increasing gradient of ACN from 0 to 30–35 % over 1–4 h.

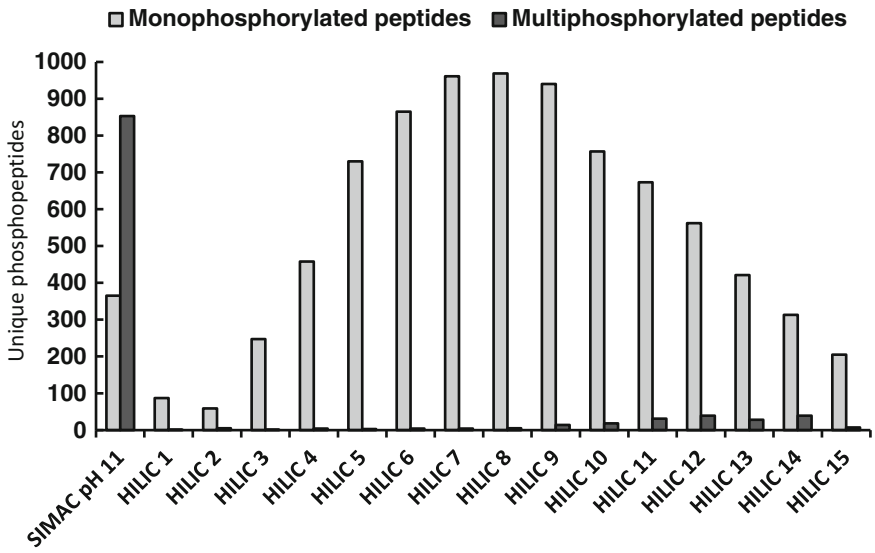


Fig. 3 Anticipated result of the TiSH enrichment-fractionation strategy after LC-MS/MS of the phosphopeptide samples. Overall, the phosphopeptide enrichment specificity is around 95 %. About 70 % of the phosphopeptides identified in the SIMAC pH 11 are multi-phosphorylated, while 77 % of the multi-phosphorylated peptides identified in the experiment originated from this fraction

2. Perform phosphopeptide identification and quantification using an appropriate software tool to generate peak lists from the raw mass spectrometry data and search it against a species-specific protein database using a database search engine (*see* Subheading 2.6, **item 2**). The anticipated results of the enrichment/fractionation strategy after LC-MS/MS analysis and data processing/database searching are shown in Fig. 3. Most of the multi-phosphorylated peptides (77 %) are identified from the SIMAC pH 11 fraction, with only 3 % of the phosphopeptides identified in the HILIC fractions being multi-phosphorylated. The multi-phosphopeptide specificity of the SIMAC pH 11 fraction is 70 %, with most mono-phosphorylated peptides identified in this fraction being very acidic, causing them to be retained to a similar degree as the multi-phosphorylated peptides. Overall, a phosphopeptide enrichment specificity of 95 % can be expected.

4 Notes

1. Triethyl ammonium bicarbonate should be employed instead of the more commonly used ammonium bicarbonate (ABC) as IMAC is sensitive to ammonium ions (Larsen MR, unpublished results), and TEAB is more volatile than ABC.

Furthermore, in contrast to TEAB, ABC will interfere with the amine-reactive chemical labeling utilized for relative quantification via stable isotope labeling.

2. Amino acid composition analysis after protein acid hydrolysis is the most accurate protein concentration measurement method [20], but requires specialized equipment and might not be available to most researchers. An alternative method requiring only a spectrophotometer is the Micro BCA Protein Assay Kit. However, it is important to ensure that the assay used is compatible with the components of the lysis buffer or at least diluted to a compatible concentration prior to the assay. Furthermore, the protein concentration reference standards should be diluted in the lysis buffer to include the same concentration of potential interfering compounds contained in the sample.
3. A recent study [21] has shown TrypZean—a recombinant bovine trypsin expressed in corn—to be a very cost-effective and specific alternative to the commonly used sequencing/mass spectrometry grade trypsin.
4. Degradation over time or batch-to-batch variation of the IMAC resin can change the affinity of the resin towards phosphopeptides. It has been reported that changing the monophosphopeptide Elution Buffer to 20 % ACN, 2 % TFA is required to achieve a specific multiphosphopeptide enrichment (unpublished results, K. Engholm-Keller (Children's Medical Research Institute, Sydney, Australia), A. Liberski and J. Graumann (Weill Cornell Medical College in Qatar, Doha, Qatar)). Optimally the enrichment should be tested using 1, 1.5 and 2 % TFA prior to performing the experiment.
5. Primary amines in proteins (N-terminals and lysines) are prone to carbamylation by isocyanate, which urea is in equilibrium with in aqueous solutions. This process is temperature dependent and while reduction and tryptic digestion is usually performed at 56 °C and 37 °C, respectively, carbamylation is reduced to low levels at room temperature [22] without significantly diminishing the efficiency of the reactions. Alternatively, to avoid carbamylation, 6 M guanidine hydrochloride can be used as a denaturing agent instead of urea [22].
6. When performing quantitative phosphoproteomics, changes in protein level due to protein expression, subcellular location etc. can affect the relative levels of phosphorylated peptides quantified by LC-MS/MS after phosphopeptide enrichment. For samples in which such protein changes occur, phospho-level normalization based on protein level changes is necessary.

7. Keep the total volume of the sample suspended in Loading Buffer at a level proportional to the sample amount (a few mL for 1–2 mg of peptide). Very diluted samples will require longer loading time than specified in this protocol.
8. The centrifugation is performed at 14,000 × *g* as the TiO_2 resin is harder to pellet in aqueous buffers due to the surface tension/viscosity.
9. Contaminating resin in the eluate can reduce phosphopeptide recovery as the phosphopeptides can re-adsorb to the resin when the eluate is acidified prior to SIMAC or RP desalting. To check for potential resin contamination of the eluate, the tube can be centrifuged for 1 min at 14,000 × *g*—if any pellet is visible, transfer the liquid to another tube.
10. The TiO_2 resin from the pre-enrichment steps (Subheading 3.2, **steps 2** and **3**) can be regenerated and reused in the later TiO_2 enrichment steps: Incubate the resin in 50 % ACN, 0.1 % TFA for 10 min after elution, centrifuge the beads at 2000 × *g* for 15 s and remove the supernatant.
11. The sample is dried to remove ammonia, which can interfere with the IMAC-based phosphopeptide enrichment in SIMAC.
12. Ensure that no IMAC resin is present in the flow-through. This can be done by loading the flow-through onto the column again, collecting the flow-through. For the eluate beads can be removed by centrifugation at 14,000 × *g* for 1 min followed by transfer of the supernatant to a new tube if an IMAC resin pellet appears.
13. Performing the 20 % ACN, 1 % TFA elution too fast will lead to mono-phosphorylated peptides in the pH 11 fraction as these species will not have been fully washed off prior to the pH 11 elution—too slowly and multiphosphorylated peptides can be eluted off the column, reducing the recovery in the pH 11 fraction.
14. In our original protocol, the phosphopeptide sample was dissolved sequentially in 10 % TFA, H_2O and finally ACN and not dissolved in DMSO prior to adjustment to HILIC solvent B conditions. However, including DMSO has later been showed to improve the solubility of large phosphopeptides and N-linked glycopeptides (unpublished results, M.R. Larsen (University of Southern Denmark)). Directly trying to dissolve the sample in 90 % ACN, 0.1 % TFA will lead to large losses of phosphopeptides not being solubilized [12].
15. The flow-trough from the column during loading most often does not contain phosphopeptides and can usually be discarded.

Acknowledgements

This work was supported by the Lundbeck Foundation (M.R.L., Junior Group Leader Fellowship, K.E.-K., postdoctoral fellowship), The Danish Council for Independent Research and the European Union FP7 Marie Curie Actions—COFUND programme (K.E.-K., MOBILEX postdoc fellowship, grant ID DFF–1325-00154) and the Sehested Hansen Foundation (M.R.L., K.E.-K.).

References

1. Neville DC, Rozanas CR, Price EM, Gruis DB, Verkman AS, Townsend RR (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. Protein Sci 6(11):2436–2445
2. Posewitz MC, Tempst P (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. Anal Chem 71(14):2883–2892
3. Ikeguchi Y, Nakamura H (1997) Determination of organic phosphates by column-switching high performance anion-exchange chromatography using on-line preconcentration on titania. Anal Sci 13(3):479–483
4. Kuroda I, Shintani Y, Motokawa M, Abe S, Furuno M (2004) Phosphopeptide-selective column-switching RP-HPLC with a titania precolumn. Anal Sci 20(9):1313–1319
5. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4(7):873–886
6. Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76(14):3935–3943
7. Sano A, Nakamura H (2004) Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. Anal Sci 20(3):565–566
8. Jensen SS, Larsen MR (2007) Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. Rapid Commun Mass Spectrom 21(22):3635–3645
9. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. Mol Cell Proteomics 7(4):661–671
10. McNulty DE, Annan RS (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. Mol Cell Proteomics 7(5):971–980
11. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 77(19):6426–6434
12. Engholm-Keller K, Birck P, Storling J, Pociot F, Mandrup-Poulsen T, Larsen MR (2012) TiSH--a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. J Proteomics 75(18):5749–5761
13. Engholm-Keller K, Hansen TA, Palmisano G, Larsen MR (2011) Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO(2) chromatography applied to proximal EGF signaling. J Proteome Res 10(12):5383–5397
14. Li QR, Ning ZB, Tang JS, Nie S, Zeng R (2009) Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. J Proteome Res 8(11):5375–5381
15. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
16. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N et al (2010) A guided tour of the trans-proteomic pipeline. Proteomics 10(6):1150–1159
17. Hsu JL, Huang SY, Chow NH, Chen SH (2003) Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem 75(24):6843–6852

18. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S et al (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3(12): 1154–1169
19. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G et al (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75(8):1895–1904
20. Noble JE, Knight AE, Reason AJ, Di Matola A, Bailey MJ (2007) A comparison of protein quantitation assays for biopharmaceutical applications. Mol Biotechnol 37(2):99–111
21. Bunkenborg J, Espadas G, Molina H (2013) Cutting edge proteomics: benchmarking of six commercial trypsins. J Proteome Res 12(8): 3631–3641
22. Kollipara L, Zahedi RP (2013) Protein carbamylation: in vivo modification or in vitro artefact? Proteomics 13(6):941–944

Chapter 12

Offline High pH Reversed-Phase Peptide Fractionation for Deep Phosphoproteome Coverage

Tanveer S. Batth and Jesper V. Olsen

Abstract

Protein phosphorylation, a process in which kinases modify serines, threonines, and tyrosines with phosphoryl groups is of major importance in eukaryotic biology. Protein phosphorylation events are key initiators of signaling responses which determine cellular outcomes after environmental and metabolic stimuli, and are thus highly regulated. Therefore, studying the mechanism of regulation by phosphorylation, and pinpointing the exact site of phosphorylation on proteins is of high importance. This protocol describes in detail a phosphoproteomics workflow for ultra-deep coverage by fractionating peptide mixtures based on high pH (basic) reversed-phase chromatography prior to phosphopeptide enrichment and mass spectrometric analysis. Peptides are separated on a C_{18} reversed-phase column under basic conditions and fractions collected in timed intervals followed by concatenation of the fractions. Each Fraction is subsequently enriched for phosphopeptides using TiO_2 followed by LC/MS analysis.

Key words Phosphoproteomics, High pH, Reversed-phase, Phosphorylation, Phosphopeptide enrichment, Fractionation

1 Introduction

Protein phosphorylation is a pivotal post-translational modification involved in regulating virtually all cellular processes. Site-specific phosphorylation often functions as a cellular switch, which can alter the activity of a protein, change its subcellular localization or interaction partners, as well as mark proteins for degradation [1]. The temporal and spatial distribution of protein phosphorylation in a cell is controlled by the action of protein kinases and counteracted by protein phosphatases in an interplay that is tightly regulated in healthy cells under physiological conditions [2–4]. However, misregulation of this cellular control as a result of genomic mutations, DNA damage, and other factors often leads to diseases such as cancer [4]. Understanding the regulation and dynamics of protein phosphorylation in health and disease is therefore of great importance. Although many different analytical

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_12,

strategies exist for investigation of the phosphoproteome, enrichment and fractionation is usually required for deep coverage. This is due to the high dynamic range and complexity of eukaryotic phosphoproteomes as well as generally low stoichiometry of phosphorylation sites and hence low abundance of phosphopeptides in total cell digests. We recently described an offline peptide fractionation method as a powerful tool to improve depth and coverage in phosphoproteome analyses [5]. The method relies on fractionating tryptic peptides on a hydrophobic C_{18} column under basic conditions (pH > 8). Although several methods and strategies exist for fractionating peptides, high pH reversed-phase fractionation has shown great promise in recent years due to its high separation and resolving power compared to standard methods such as those based on ion exchange or hydrophilic interaction chromatography [6]. Fractionation based on high pH reversed-phase requires concatenation of fractions from different parts of the elution gradient to be orthogonal with downstream low pH reversed-phase separation in tandem with mass spectrometry analysis [7]. Following collection and concatenation of fractions, phosphopeptides are enriched using titanium dioxide beads [8]. Depending on starting amounts, it is possible to map close to 40,000 phosphopeptides within 24 h of mass spectrometric analysis time using this technique.

2 Materials

All buffers are made with sequencing grade chemicals and ultrapure water (Milli-Q).

2.1 Cell Lysis and Digestion

1. Eukaryotic cells (adherent or suspension) grown in large amounts (*see* **Note 1**).
2. Cell lysis buffer such as RIPA, urea, or guanidinium hydrochloride (*see* **Note 2**).
3. Acetone.
4. Bradford or other protein concentration determination assay.
5. Reducing reagents such as tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT) (*see* **Note 3**).
6. Alkylation reagent chloroacetamide (CAA).
7. Lys-C and trypsin.

2.2 Buffers for Fractionation

1. Tryptic digested peptide mixture in basic condition (*see* **Note 4**).
2. BSA or other peptide standard mixture.
3. Buffer A, 5 mM ammonium hydroxide (*see* **Note 5**).
4. Buffer B, 90 % acetonitrile, 5 mM ammonium hydroxide.
5. Optional Buffer C for compatible LC systems (*see* **Note 6**).

2.3 Fractionation and Concatenation

1. High pH Compatible C_{18} reversed-phase column, 4.6 mm ID × 250 mm (*see* **Note 7**).
2. High performance liquid chromatography (HPLC) system compatible up to 400 bars (5800 psi) with auto sampler, UV detector, pressure sensor, and fraction collector (*see* **Note 6**) is recommended.
3. 96 deep well plates.

2.4 Phosphopeptide Enrichment and STAGE Tipping

1. Titanium dioxide (TiO_2) beads.
2. 2,5-dihydroxybenzoic acid (20 mg/ml) in 80 % ACN, 1 % TFA.
3. C8 STAGE tip and C18 STAGE tip.
4. Trifluoroacetic acid (TFA).
5. 5 % ammonium hydroxide (NH_4OH).
6. 20 % ammonium hydroxide (NH_4OH), 25 % acetonitrile.

2.5 Nano-Liquid Chromatography (nLC) and Mass Spectrometer (MS)

1. Capillary (75 μm internal diameter, 15 cm length) column packed with 3 μm or smaller C_{18} beads.
2. Nano-LC system compatible up to 280 bars (for 15 cm column), coupled to MS.
3. High performance mass spectrometers capable of MS analysis at high resolution (>10,000) such as Orbitrap Q-Exactive (Thermo Fischer) or time of flight instruments such as 5600 Q-TOF (AB Sciex).

3 Methods

3.1 Cell Lysis and Protease Trypsin Digestion

1. Lyse cells using RIPA, urea or guanidinium hydrochloride (*see* **Note 2**).
2. If using RIPA buffer, acetone precipitate proteins, centrifuge to remove acetone and resuspend protein pellet in urea or guanidinium hydrochloride.
3. Determine protein concentration using Bradford or other protein assay.
4. Reduce proteins with DTT (1 mM) or TCEP (5 mM) for 30 min at RT.
5. Alkylate cysteines with CAA (5 mM) at RT in the dark. This can be done simultaneously in combination with TCEP (*see* **Note 3**).
6. Add Lys-C (1:100 protease–protein ratio) at room temperature for 4 h.
7. Dilute urea to >2 M (>1 M for guanidinium hydrochloride) with 50 mM ammonium bicarbonate buffer.
8. Add trypsin (1:50 protease–protein ratio), digest overnight at 37 °C.

3.2 Sample Preparation for Fractionation

1. Tryptic peptide mixture should be prepared and desalted on reversed-phase C_{18} cartridges such as C_{18} sep-pak to remove salts prior to fractionation. Cell lysis additives (such as detergents), which can interfere with fractionation and other downstream analysis must be removed.
2. Speedvac peptide mixture to evaporate acetonitrile (*see* **Note 8**).
3. Add Buffer A to the sample to bring it to roughly same starting condition (i.e., pH and mobile phase additives such as ammonium hydroxide similar to running buffers) as LC. Sample volume should be kept below the injection maximum of the LC system autosampler and injection loop.
4. Prepare BSA peptide mixture in a similar fashion.

3.3 High Pressure Liquid Chromatography Preparation (HPLC)

1. Prepare column by connecting to HPLC system.
2. Depending on the internal diameter of the column and manufacturer instructions, run Buffer A with recommended optimal flow rate (*see* **Notes 6** and **9**).
3. Monitor peptide elution by UV at absorbance 214 nm wavelength, which is specific for peptide bonds. Absorbance at 260 and 280 nm may also be used however they are not as accurate.
4. Monitor column pressure over time so that it reaches a stable baseline (*see* **Note 9**).
5. Maintain flow rate on the column until a stable UV baseline is observed and expected column pressure is achieved. This may take several minutes depending on column condition and equilibration (*see* **Note 9**).

3.4 Gradient Design

Please note that elevated column temperatures for separation at high pH are detrimental for column bed material. We therefore recommend performing fractionation at ambient temperatures (~25 °C). *See* Fig. 1 for example of typical gradient used for fractionation.

1. Design a gradient starting from low Buffer B composition (0–5 %) to 25 % in 60–90 min (*see* **Note 10**).
2. The next part of the gradient should go from 25 to 40 % Buffer B in 5–10 min.
3. Add a quick ramp up to 60–70 % B in 5 min. At this point the fraction collection can be stopped.
4. Maintain gradient at high Buffer B for up to 5 min before ramping back down to 0–5 % Buffer B.
5. Hold at low Buffer B for 5–10 min to allow for column re-equilibration.

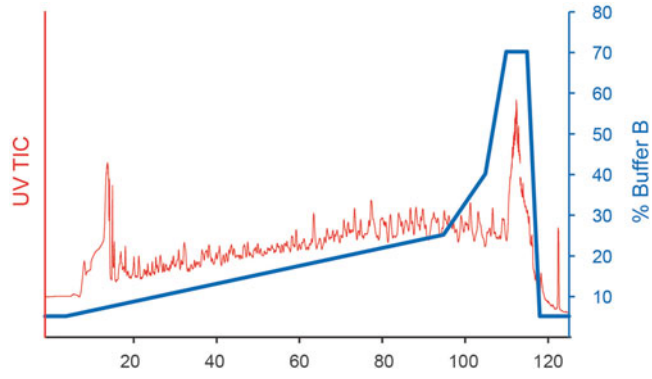


Fig. 1 Fractionation gradient of Buffer B concentration is overlapped with the UV total ion chromatogram (TIC)

3.5 Determining Column Condition

Before and after fractionating peptide mixtures at high pH, it is recommended to determine the condition of the column using a standardized peptide mixture. Steps for determining column condition and separation capacity using bovine albumin (BSA) peptide mixture are recommended below. Each run should be monitored using UV absorbance and pressure. BSA sample does not need to be fraction collected and can be sent to waste.

1. Design gradient using buffers without ammonium hydroxide for analysis of BSA peptides. Start from 5 % B to 30 % B in 20–30 min.
2. Ramp B from 30 to 40 % in 5 min.
3. Ramp B from 40 to 60 % in 2 min, hold for 2–5 min.
4. Decrease B from 60 to 5 % in 2 min.
5. Maintain at 5 % B for 5–10 min.
6. Inject 10–100 μg of BSA peptides onto column using described gradient.
7. Determine baseline and peak conditions using UV trace (*see* **Note 11**).

3.6 Sample Injection and Fraction Collection

Please note that it is recommended to first run a BSA mixture prior to running fractionating on a new column (*see* Subheading 3.5).

1. Prepare sample for injection onto the column using autosampler or with an additional line (*see* **Note 12**). Inject volume based on sample concentration and column capacity (roughly 5–8 mg for 4.6 mm ID × 250 mm C_{18} columns).
2. Run program with gradient described in Subheading 3.3.

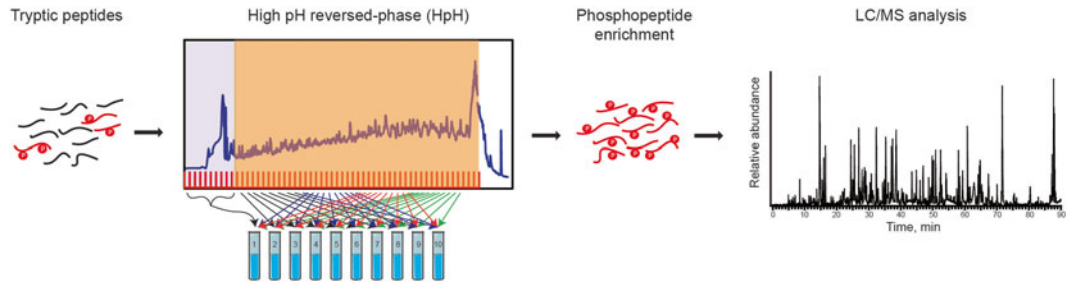


Fig. 2 Basic workflow for fractionation with high pH reversed-phase is presented. Tryptic peptides are fractionated on a reversed-phase C_{18} column using high pH buffers. The fractions are then concatenated to mix different parts of the gradient so they are orthogonal to downstream LC/MS analysis at low pH. Fractions are then enriched for phosphopeptides and analyzed via LC/MS

3. Collect fractions every 30–60 s in a deep well plate or Eppendorf tubes, depending on fraction collector and rack type (*see* **Note 13**).
4. Store samples at 4 °C or proceed to next step (samples can be stored for up to 2 weeks).

3.7 Fraction Concatenation

1. Prepare samples for concatenation by first pooling flow-through region as one pre-concatenated fraction (*see* Fig. 2).
2. Concatenate to 10–15 fractions (*see* Figs. 2 and 3).
3. Although not necessary some fractions such as the first fraction have larger volumes that can be decreased by drying in a speed-vac (*see* **Note 14**).

3.8 Post Fractionation Column Care

In order to extend the life of the fractionation column some steps are recommended below (*see* **Note 15**).

1. After running the fractionation gradient a post wash method is recommended with buffers containing no ammonium hydroxide and only water and acetonitrile as Buffers A and B.
2. Design a wash gradient which ramps Buffer B (acetonitrile) from 5 to 90 % in 15–20 min where it should be held for 5–10 min.
3. Ramp B down to 5 % in 5 min.
4. Repeat **steps 2** and **3**.
5. End step with 50–70 % B running at low flow rates (0.05–0.100 µl/min) where the column can sit idle. If column is not to be used for long periods of time it is recommended to cap and store the column (see next step).
6. Cap column from one side and let pressure build up slightly while low flow is still running.
7. Quickly remove column from the other side and cap column for storage.

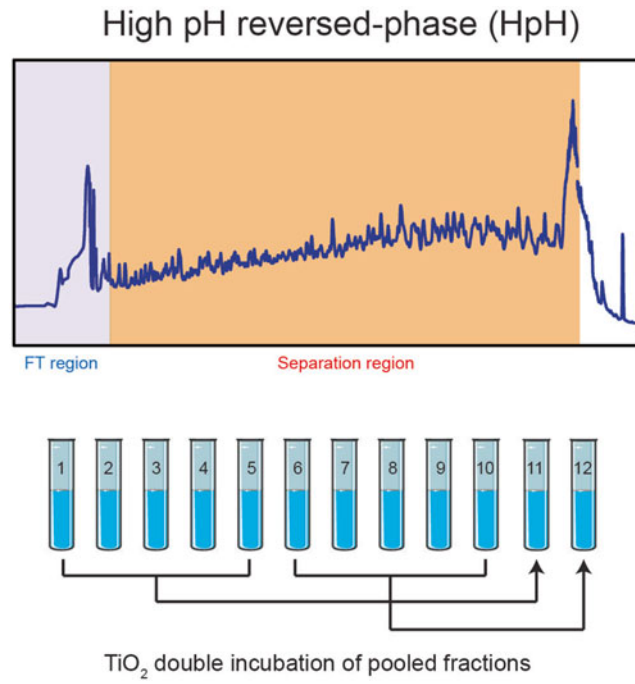


Fig. 3 Strategy for double incubation of fractions with TiO_2 beads is presented. After each fraction is enriched for phosphopeptides, the supernatants are pooled together for double TiO_2 enrichment

3.9 Preparing Fractions for Phosphopeptide Enrichment

The following steps are an outline for preparation of fractions for enrichment of phosphopeptides with titanium dioxide (TiO_2) beads.

1. Measure peptide concentration of each fraction using A_{280} nanodrop or other estimation methods compatible with peptides such as Lowry assay.
2. Bring the fractionated peptide solution to conditions compatible with enrichment method. For TiO_2 enrichment, add acetonitrile to 60–80 % and trifluoroacetic acid (TFA) to a final concentration of 5–6 %.
3. Samples may also be stored in this condition at −20 °C for several days up to a few weeks.

3.10 Enrichment of Phosphopeptides and Preparation for LC/MS Analysis

We recommend fractionating peptide mixtures prior to enrichment of phosphopeptides. The resultant reduction in sample complexity leads to more efficient enrichment of phosphopeptide mixtures and thus deeper coverage. The method described below is for enrichment using titanium dioxide beads (MOAC), please consult

manufacturer protocols for different enrichment strategies such as IMAC.

1. Weigh TiO_2 beads (1:2 peptide–bead ratio) and resuspend in DHB buffer (we recommend 10 µl DHB buffer per milligram of TiO_2 beads). Incubate for 20 min while rotating or shaking.
2. Add TiO_2 bead solution to the fractionated peptide solution at a ratio of 2 mg beads (20 µl of TiO_2—DHB solution in this case) per milligram of peptides.
3. Incubate fractions with TiO_2 beads for 30 min at room temperature on a sample rotator.
4. Centrifuge fractions to form a pellet with TiO_2 beads. Do not discard supernatant if performing proteomics analysis. Also keep the supernatant if performing another enrichment round to recover multiphosphorylated peptides as described in Subheading 3.9, **steps 6** and 7.
5. Resuspend pellet with 100 µl of 50 % acetonitrile and 5 % TFA.
6. After centrifugation pool supernatants together into 1–3 fractions (*see* Fig. 2).
7. Perform Subheading 3.9, **steps 2–5** on the pooled fractions to enrich pooled fractions again for phosphopeptides leading to a higher recovery of multiply phosphorylated peptides.
8. Load beads suspended in 100 µl of 50 % acetonitrile and 5 % TFA on C_8 STAGE tip.
9. Centrifuge 3 min at 800×*g* to remove buffer followed by washing with 30 % acetonitrile, 1 % TFA (add additional centrifugation time, if not all the buffer has washed through the STAGE tip).
10. Centrifuge to remove buffer followed by washing with 60 % acetonitrile with 1 % TFA. Elute with 20 µl 5 % NH_4OH, followed by 20 µl of 20 % NH_4OH with 25 % acetonitrile into a new tube.
11. Evaporate (not to dryness) samples in speedvac for 15 min at 45 °C.
12. Resuspend samples in acidic buffer (1 % TFA, 2 % acetonitrile).
13. Prepare C18 STAGE tip by washing/centrifuging for 2 min at 800×*g* each with methanol (20 µl), 80 % acetonitrile with 0.1 % TFA (20 µl), 3 % acetonitrile, 0.1 % TFA (40 µl). Last wash might require additional centrifugation time.
14. Load samples on C18 STAGE tip, centrifuge at 800×*g* until the sample is washed through and wash with 3 % acetonitrile with 1 % TFA (samples can be stored in this condition).
15. Elute samples with buffers containing 40–60 % acetonitrile, 0.1 % TFA. Speedvac and resuspend in Buffer A.

3.11 LC/MS

Prior to LC/MS analysis it is important to have a method ready for phosphoproteome analysis. The details described below are for a Q-Exactive mass spectrometer coupled to a nanoLC system. For different systems we recommend consulting manufacturer settings or published literature for guidance. For phosphopeptide analysis, we have modified the method for sensitive analysis; this is due to the fact that sample complexity is reduced in fractionated and phospho-enriched samples where quality of the fragment HCD (high collisional dissociation) spectrum is favored over sequencing speed [9].

1. Design a LC method with nLC gradient of 60 min or more. If not limited by instrument time we recommend 120 min.
2. Include ramp up to 80 % Buffer B (80 % acetonitrile) and back down to 5 % B (starting conditions of LC method).
3. Injection volume should be kept to 5 µl or below for each sample (*see* **Note 16**).
4. Design a top 10 method on the Q-Exactive with 70,000 MS resolution and 35,000 dd-ms2 resolution.
5. The scan range for the MS should be set to 400–1600 m/z, for dd-ms2 fixed first mass should be set to 100 m/z.
6. Maximum injection time for the MS should be 20 ms, injection time for dd-ms2 should be 108 ms so it is in parallel with the Orbitrap transient at 35,000 resolution.
7. Automatic gain control AGC target for MS should be set to 1e6 or higher, for dd-ms2 it should be 10× lower (100,000 in this case).
8. Quadrupole isolation window (width) should be set to 1 m/z (but not lower)—2 m/z (not higher).
9. Normalized collisional energy should be set to a value between 25 and 30 (should be higher for phosphopeptides), we recommend 28 as safe median.
10. Dynamic exclusion should be set to 30 s.
11. Run samples with this method.

3.12 Data Analysis

After raw files have been collected, use LC/MS data analysis software of your choice. It is recommended to use software packages which can perform data analysis for post-translational modifications. Software packages such as MaxQuant (Mann Lab, www.maxquant.org), Mascot (Matrix Science), and Proteome Discoverer (Thermo Scientific) are good options. When performing analysis with the software package it is important to specify phosphorylation modification of serine, threonine, and tyrosine as variable modifications. Additionally, we also recommend the addition of protein acetyl-n terminus, methionine oxidation as variable modification. Since cysteines are modified to completion using

alkylating reagent (CAA in our case) it is important to set that as a fixed modification (carbamidomethylation for CAA). If using a different alkylating reagent, it is important to specify the correct fixed modification. When using high resolution mass spectrometers, it is important to limit the parts per million tolerance at the MS and MS/MS levels in order to increase the confidence for peptide hits. This is important for PTM analysis where proper localization and peptide sequence coverage of the modification is very important for site localization. When searching against the FASTA database of your organism, it is crucial that peptide false discovery rate is set to 1 % or lower. Vendor instructions for each software packages as well as published literature can aid in selecting proper parameters for each software package [10].

4 Notes

1. The shown experiments were done with tryptic digests of whole cell lysates from mammalian HeLa cervix carcinoma cells. Cells were serum starved overnight in DMEM and stimulated for 10 min with serum prior to cell lysis.
2. There are various options available for lysing cells. Lysis buffer containing detergents such as SDS require additional steps for removal prior to LC/MS analysis. We recommend guanidinium hydrochloride in place of urea as it does not lead to additional peptide modifications (such as carbamylation) even at elevated temperatures. For the purpose of this chapter, workflow based on RIPA based lysis is presented.
3. We recommend TCEP instead of DTT as reducing agent since it allows for the reduction and alkylation step (with CAA) to be done simultaneously. Additionally, TCEP lacks a strong odor unlike DTT (*see* Chapter 17 for more details).
4. It is important to have peptide mixtures in buffer conditions similar to starting condition of the LC gradient. For example if the gradient starts at pH 10 with buffer A, it is recommended to reconstitute the peptide mixture in Buffer A or add Buffer A to the sample so the pH is similar. This is to allow for proper peptide hydrophobic interaction in similar conditions as the column.
5. Addition of buffers such as ammonium formate is not recommended since they have been shown to cause column degradation in combination with high pH [5, 11]. At 5 mM ammonium hydroxide the pH value is roughly 10.5, due to the absence of any buffer ions the pH value does not drop or increase much in the range of 1–5 mM. Ammonia dissolved up to 25 % will have a concentration of roughly 14–14.5 N. We recommend monitoring pH with pH meter or strips, buffers should be exchanged every few weeks.

6. For HPLC systems with quaternary pumps (systems in which four different lines can be operated simultaneously), we recommend allocating line A to pure water (Milli-Q grade), line B to pure acetonitrile, and line C to higher concentration of ammonium hydroxide (10–50 mM). In this case line C would be running constantly at 10 % of total flow rate to dilute the concentration of the ammonium hydroxide to a final concentration of 1–5 mM. These systems are recommended as line C can be shut off after the fraction collection period and the column washed without ammonium hydroxide.
7. It is important to determine buffer compatibility of columns with manufacturers. We recommend columns with small bead sizes (<5 μM) as they better separate peptides with small peak widths. Columns with smaller internal diameters (<4.6 mm) will require smaller loading amounts of tryptic peptides. In all cases we recommend column lengths of 250 mm since that is sufficient for efficient separation using long gradients. Columns with bigger internal diameters will require the HPLC to handle higher backpressure. We also recommend addition of a guard column with guard cartridges as they can extend the lifespan of the column without affecting separation. The principle of the guard is that the sample is loaded onto the guard column and separated on the main column during the gradient. This prevents larger moieties from affecting the main column and reducing the amount of particulates which may accumulate on the main column. This increases the lifespan of the main separation analytical column since guard cartridges are easily replaceable. This solution is furthermore economically favorable as the cost-benefit of including a replaceable guard column before the separation column can prevent frequent replacement of the analytical column, which is manyfold more expensive than guard column cartridges.
8. It is important that acetonitrile in the sample is removed prior to injection onto the column as it can prevent efficient hydrophobic interactions and broaden peaks as they elute. Typically 30–60 min of drying at 45–60 °C is sufficient to remove acetonitrile. This can also be assessed by eye especially if peptides were eluted off a Sep-Pak in 50 % acetonitrile. The final volume after drying in a speedvac should be half or less of the starting volume since acetonitrile is more volatile compared to water and will dry off faster. It is possible to dry samples for longer periods of time or even to dryness and resuspend in starting buffer with desired volume (we do not however, recommend this since reconstituting peptides can lead to loss of certain peptides and highly acidic buffers are required to resuspend dried peptides, thus making the starting conditions incompatible with high pH, therefore we suggest drying to low volumes but not dryness).

9. For 4.6 mm ID × 250 mm C_{18} columns flow rates of 1 ml/min are recommended. Typical pressure range when running at 1 ml/min (with low organic) is 150–250 bars depending on bead material. Consult column manufacturer guidelines for further assistance. Instructions for flow rates are usually available from column manufacturer which we highly recommend consulting before operating the HPLC with column. Typically manufactured columns are shipped capped with high organic buffer in the column. It is important that the user recognizes this and washes the column with starting condition of low Buffer B and high Buffer A for several minutes and column volumes. Additionally monitoring the UV and pressure baseline can help determine when the column has reached equilibration as the column has lower backpressure with higher organic solvent. Running the column with several short empty runs with a quick gradient to high organic and back to low organic can help acclimate the column to running buffers for peptide separation. We do not recommend fractionating a peptide mixture on a new column for the first run.
10. Phosphopeptides typically elute earlier in the acetonitrile gradient compared to normal peptides at high pH due to phosphate groups carrying negative charges and preventing hydrophobic interactions with the column [12]. Therefore most of the separation is likely to occur earlier in the gradient (up to 20 % acetonitrile) and thus the gradient length should reflect this when separating phosphopeptides.
11. It is recommended to run BSA peptide mixtures using this method before running any fractions on the column. Advantages of this are to determine the column deterioration and condition since fractionation at high pH may adversely affect the separation efficiency of the column [5, 11]. UV trace of the BSA peptides should be sharp, symmetrical, and the peak width should be small. Several peaks should also be separated across the gradient giving the user an idea of the hydrophobic separation. Under ideal conditions, the UV trace of peptide mixtures should not deviate much from the first few BSA runs and after fractionation of several samples. By having a standard to base the condition of the column, determining column lifespan as well as discarding the column when unfit for proper separations is possible.
12. If the available injection loop of the auto-sampler is small and limits the injection volume other methods of injection of the sample onto the column should be utilized. We recommend injection loops of 500 µl or more. In cases where a larger loop is not present or the final volume is too high, it is possible to inject with an extra line or pump if present. However this runs the risk of contaminating your HPLC system with sample and is not recommended.

13. 96 deep well plates are recommended as they hold 2 ml per well and allow for longer gradient times for collection. Most fraction racks for Eppendorf or other tubes are limited in the number of tubes they may hold. It is recommended to select correct settings for rack type and tubes using the LC control software.
14. It might be necessary to speedvac fractions if performing proteomics analysis in which small amounts can be taken out (i.e., if 2 mg of peptides are fractionated in ten fractions with basic reversed-phase, one can expect 200 μg per fraction, from which only 1 % (2 μg) of the fraction would theoretically be required for proteomics analysis since the injection amounts for nano-LC columns coupled to MS are typically within this range). It is recommended to determine the concentration using Nanodrop A_{280}. Depending on gradient length and number of fractions, the fraction volume can vary from 5 to 15 ml.
15. Monitor wash method with UV and pressure changes, UV absorbance counts should be low relative to UV trace of previous sample. This can be helpful in determining the amount of residual sample if some persists.
16. The higher the injection volume, the longer the loading of the sample onto the nano column takes. For a typical 15 cm nano column (precolumn) on a standard LC (maximum 300 bars), injecting 5 μl of sample at maximum pressure can take from 20 to 30 min. This is reduced with nLC systems which are UPLC (high pressure compatible LCs) which can tolerate up to 1000 bars.

Acknowledgements

The authors would like to thank members of the Proteomics Program at the Novo Nordisk Foundation Center for Protein Research (CPR) for critical input on the protocol. Work at CPR is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001).

References

1. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141: 1117–1134
2. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635–648
3. Kruse J-P, Gu W (2009) Modes of p53 regulation. Cell 137:609–622
4. Citri A, Yarden Y (2006) EGF–ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol 7:505–516
5. Batth TS, Francavilla C, Olsen JV (2014) Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. J Proteome Res 13:6176
6. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Two-dimensional separation of peptides using

RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci 28:1694–1703

7. Yang F, Shen Y, Camp DG, Smith RD (2012) High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. Expert Rev Proteomics 9:129–134
8. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76(14):3935–3943
9. Kelstrup CD, Young C, Lavallee R, Nielsen ML, Olsen JV (2012) Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. J Proteome Res 11:3487
10. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
11. Percy AJ, Simon R, Chambers AG, Borchers CH (2014) Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation. J Proteomics 106:113–124
12. Ficarro SB, Zhang Y, Carrasco-Alfonso MJ, Garg B, Adelmant G, Webber JT, Luckey CJ, Marto JA (2011) Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. Mol Cell Proteomics 10:O111.011064

Chapter 13

Phosphopeptide Enrichment Using Various Magnetic Nanocomposites: An Overview

Íris L. Batalha and Ana Cecília A. Roque

Abstract

Magnetic nanocomposites are hybrid structures consisting of an iron oxide (Fe_3O_4/γ-Fe_2O_3) superparamagnetic core and a coating shell which presents affinity for a specific target molecule. Within the scope of phosphopeptide enrichment, the magnetic core is usually first functionalized with an intermediate layer of silica or carbon to improve dispersibility and increase specific area, and then with an outer layer of a phosphate-affinity material. Fe_3O_4-coating materials include metal oxides, rare earth metal-based compounds, immobilized-metal ions, polymers, and many others. This chapter provides a generic overview of the different materials that can be found in literature and their advantages and drawbacks.

Key words Magnetic nanocomposites, Phosphopeptide enrichment, Affinity, Hybrid materials

1 Introduction

Phosphorylation is a dynamic, abundant, and highly studied post-translational modification of proteins, affecting approximately one-third of all proteins at any particular time [1]. It is involved in the regulation of a variety of cellular processes, such as cell cycle control, DNA damage responses, transcription, protein trafficking, metabolism, and programmed cell death [2, 1]. In the human proteome, phosphorylation occurs predominantly in serine (≈90 %), followed by threonine (≈10 %) and tyrosine (≈0.05 %), although the relative abundances of these residues depend on the methodology used for the quantification of phosphorylation events [3, 4].

Phosphorylation events are key players in cellular signaling, being involved in the occurrence of many human diseases, such as cancer [5], Alzheimer's [6], Parkinson's [7], cardiovascular diseases [8], schizophrenia [9], and many others. The accurate identification of phosphorylated proteins, determination of their phosphorylation sites and quantification of stoichiometry, and monitoring of temporal dynamics of protein phosphorylation in response to cellular perturbations, are fundamental to understand

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_13,

the mechanisms behind disease pathologies and inspire the development of novel biomarkers and therapeutic agents [10].

During the last decades, mass spectrometry (MS) has been the basis of enormous scientific breakthroughs and it has been playing a central role in the profiling of protein phosphorylation [3, 11, 12]. However, as phosphopeptides have comparably low ionization efficiencies and are usually present in sub-stoichiometric concentrations in biological samples, an enrichment step is required prior to MS analysis [13, 14].

A great number of enrichment methods have been reported in recent years, namely immunoprecipitation, immobilized metal affinity chromatography (IMAC), metal oxide affinity chromatography (MOAC), ion-exchange chromatography, chemical tagging and use of phosphate-affinity ligands [15].

Ion-exchange methodologies (strong ion exchange (SCX) or strong anion exchange (SAX)) usually require high amounts of starting material and prolonged analysis times, due to the presence of multiple fractions [10]. In addition, due to the nonspecific adsorption behavior of SCX and SAX, these methods are more commonly used as pre-fractionation techniques in combination with other methods such as reversed phase-LC, IMAC or MOAC [16–18]. Recently, a mixed-bed resin comprising a blend of anion and cation exchangers (ACE) has been reported to increase phosphopeptide identification by 94 % when compared to SCX [19].

Methods based on the chemical modification of the phosphate group are usually labor-intensive and implicate several reaction steps, which makes them less useful for routine utilization [20].

Phospho-specific antibodies are expensive and difficult to produce due to the low immunogenicity of the phosphate group and its susceptibility to cleavage during the immunization process [21]. Nonetheless, antiphosphotyrosine antibodies have been successfully used for the detection of a variety of phosphorylated proteins, in contrast with antiphosphoserine and antiphosphothreonine antibodies, which are still far less specific [10, 15, 22].

Metal chelating methodologies include IMAC, MOAC, and some new metal chelating ligands (e.g., Phos-Tag™) (*see* Chapter 3). These methods are based on the coordination of the negatively charged phosphate group to positively charged metal ions, a main disadvantage being the nonspecific binding of acidic peptides to the metal-chelating resins [22]. Different research groups have been trying to circumvent this problem. *O*-methyl esterification of the carboxylic acid groups is a common approach, but has some limitations associated with the occurrence of side reactions and incomplete derivatization of the carboxylic groups [23–25]. Larsen and coworkers introduced a competitive binder—2,5-dihydroxy benzoic acid (2,5-DHB)—in buffers to reduce the nonspecific binding of acidic peptides to TiO_2 microcolumns [25]. Ye and coworkers developed an optimized Fe(III) nitrilotriacetic acid (NTA) IMAC

protocol by using 60 % (v/v) acetonitrile (ACN) in loading and washing buffers, since ACN has different effects on the degree of ionization of phosphate and carboxylic acid moieties [26].

Hybrid materials, which combine distinct properties of various materials, have been proven exquisitely efficient in the selective binding of phosphorylated peptides. In particular, composite materials in the nanomolar range (nanocomposites) present higher surface area to volume ratios, providing superior surface functionalization [27, 28]. This chapter will explore the application of different classes of nanocomposites in the phosphoproteomics field.

2 Magnetic Nanocomposites

As mentioned previously, metal oxides can be used to enrich phosphorylated species in peptide samples. Among them, iron oxide nanoparticles (MNPs) present unique properties, such as superparamagnetism, high magnetic susceptibility, high coercivity, and low Curie temperature [29]. Bare MNPs have been reported to selectively bind phosphorylated peptides from tryptic digests containing 1 pmol β-casein, cytochrome c, bovine serum albumin (BSA), and horse heart myoglobin. However, this was only observed when the iron oxide was in the form of magnetite (Fe_3O_4), as maghemite (γ-Fe_2O_3) beads did not facilitate phosphopeptide binding. The reasons behind the superior performance of Fe_3O_4 over γ-Fe_2O_3 are not yet fully understood, but have been speculated to be related to structural differences and magnetic properties of those materials [30].

In addition to their easy manipulation by an external magnetic field, their surface can be functionalized with a variety of organic and inorganic materials, which not only protect them against oxidation and erosion by acids and bases, but also tailor their surface in terms of charge, hydrophobicity, and chemical functionality [31, 32]. The process of phosphopeptide enrichment using magnetic nanocomposites is illustrated in Fig. 1.

2.1 Metal Oxide-Based Magnetic Nanocomposites

2.1.1 Titanium Dioxide (TiO_2)

TiO_2 was the first metal oxide to be reported for the enrichment of phosphorylated peptides. TiO_2-based materials are commercially available and TiO_2-based protocols are common practice in many laboratories, due to their simplicity and efficiency (*see* Table 1) (*see* Chapters 9, 12, 17). Chen and Chen used a two-step sol–gel process to produce TiO_2-coated MNPs (Fe_3O_4@TiO_2) combining the convenient separation ability of MNPs with the phosphopeptide trapping capacity of TiO_2. In addition, Fe_3O_4@TiO_2 magnetic nanoparticles were effectively used as surface-assisted laser desorption/ionization (SALDI) MS matrices [33]. However, they obtained composite materials of ill-defined structure, which compromised their bias towards phosphorylated peptides, as nonspecific binding

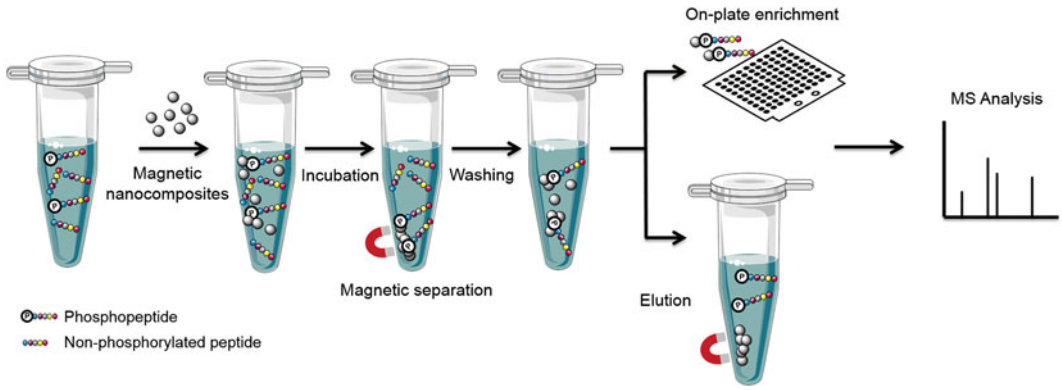


Fig. 1 Enrichment of phosphorylated peptides using magnetic nanocomposites: (1) incubation of magnetic nanocomposites with tryptic protein digest samples; (2) magnetic separation of phosphopeptide-bound magnetic nanocomposites; (3) washing step to remove nonspecifically adsorbed non-phosphorylated peptides; (4a) phosphopeptide-bound magnetic nanocomposites can be spotted onto a MALDI target and directly analyzed by MS or (4b) phosphopeptides may be eluted from particles and analyzed either by MALDI or ESI MS

of acidic peptides was also observed. Later on, Li and coworkers addressed this issue by coating the particles with a carbon layer in an intermediate reaction step in order to reduce nonspecific adsorption. Their three-step synthesis route consisted of: (1) synthesis of MNPs via solvothermal reaction; (2) MNPs coating with a thin layer of carbon (≈20 nm thickness); and (3) absorption of nanosized titanium oligomers and conversion into titanium by calcination [34, 35].

In 2010, Lu and coworkers reported the synthesis of self-assembled mesoporous TiO_2 nanocrystal structures with high adsorption capacity, low detection limit (10 fmol for a β-casein tryptic digest), high selectivity, high water dispersibility, and high chemical and mechanical stability. The process of fabrication of these mesoporous nanocrystal clusters consists of the synthesis and self-assembly of TiO_2 nanocrystals; coating with a thin layer of silica to avoid aggregation; calcination at high temperatures to improve mechanical stability and remove organic surfactants; and silica removal by etching. One of the advantages of this self-assembly process is that multiple components may be added to the clusters, which was demonstrated by the successful fabrication of γ-Fe_2O_3/TiO_2 composite clusters to facilitate separation. Both TiO_2 and γ-Fe_2O_3/TiO_2 colloidal nanocrystal clusters presented similar hydrodynamic diameters between 50 and 200 nm [36].

2.1.2 Zirconium Dioxide (ZrO_2)

A fast method using ZrO_2-coated MNPs as concentrating probes was developed to enrich samples in phosphopeptides in only 30 s employing vigorous mixing by pipetting. In addition, these particles can function as microwave absorbers and facilitate enzymatic digestion (digestion time ≈15 s sonication + 1 min microwave heating) [37].

Table 1
Commercially available magnetic nanocomposites for phosphopeptide and phosphoprotein enrichment

Affinity ligand	Kit	Highlights	Company
TiO_2	TiO_2 Mag Sepharose™	– Parallel handling of samples—processing of 6 samples in <1 h	GE Healthcare Life Sciences www.gelifesciences.com/
TiO_2	Phos-trap™	– Allows to enrich 1–96 samples at a time in <10 min	PerkinElmer www.perkinelmer.com
TiO_2	Pierce Magnetic Titanium Dioxide Phosphopeptide Enrichment Kit	– Processing of 1–96 samples in ≈15 min – Enrichment of less than 100 fmol phosphoprotein – 1000 Greater sensitivity than traditional IMAC	Thermo Scientific Pierce www.piercenet.com/
N.A.	Captivate™ Microscale Phosphopeptide Isolation Kit	– Binding capacity of 1–2 pmol/μg ferrofluid – Distinction between pSer, pThr, and pTyr due to selective β-elimination/addition modification reactions	Life Technologies www.lifetechnologies.com/
N.A.	TALON® PMAC	– Non-denaturing protocol allows phosphoprotein enrichment maintaining protein conformation and solubility – Allows enrichment from any cell or tissue sample	Clontech www.clontech.com/
Ti^{4+}-Polyamidoamine generation 4 dendrimer	PolyMAC-Ti	– Digested complex cell lysates from DG-75 B-cell lymphoma cells reveal that PolyMAC identified a higher number of phosphosites and presented higher enrichment selectivity	Tymora tymora-analytical.com/

N.A. not available

MNPs coated with both TiO_2 and ZrO_2 (Fe_3O_4@TiO_2-ZrO_2) present improved phosphopeptide trapping ability when compared to Fe_3O_4@TiO_2 or Fe_3O_4@ZrO_2 alone and efficiently enrich samples in both mono- and multi-phosphorylated peptides [38]. These metal oxides have complementary properties: ZrO_2 is more selective towards mono-phosphorylated peptides whereas TiO_2 preferentially binds to multi-phosphorylated peptides [39].

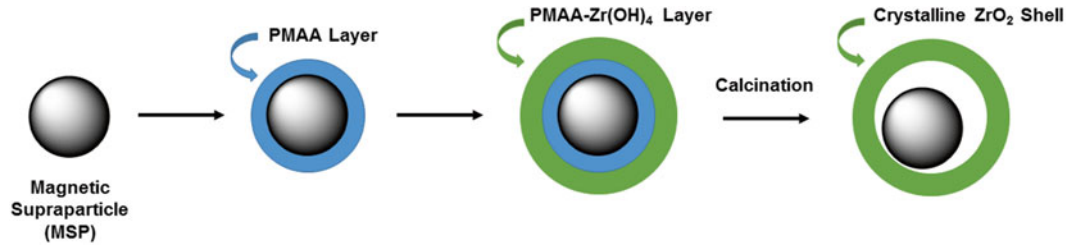


Fig. 2 Synthesis of yolk–shell MSP@ZrO_2. Magnetic supraparticles synthesized by a solvothermal reaction are modified with a polymeric layer of polymethylacrylic acid (PMAA), followed by functionalization with $Zr(OH)_4$. Calcination of MSP@PMAA@PMAA-$Zr(OH)_4$ removes the polymeric layer and leads to the formation of hollow crystalline ZrO_2 spheres with movable MSP core

Very recently, yolk-shell magnetic supraparticles coated with ZrO_2 (MSP@ZrO_2) have been reported to selectively enrich phosphopeptides both from standard phosphoprotein tryptic digests and biological samples [40]. Supraparticles are originated from the assembly of MNPs and have enhanced magnetic responsiveness and preserved superparamagnetism [41]. The yolk-shell architecture consists of nanoparticle core inside a hollow shell, presenting low mass density which improves dispersibility and enhances the adsorbing efficiency of phosphopeptides when compared to solid Fe_3O_4@ZrO_2 microspheres (*see* Fig. 2) [40]. A total of 33 phosphopeptides containing 49 phosphorylation sites mapped to 33 phosphoproteins were identified in human saliva, which was a better result when compared to zinc oxide (ZnO)-coated MNPs. Nonetheless, ZnO-coated MNPs presented a detection limit 3 orders of magnitude lower (2.5 fmol for β-casein tryptic digest) and an enrichment time 60 times lower than the yolk-shell MSP@ZrO_2 [40, 42]. In addition, TiO_2-coated MNPs presented similar phosphopeptide trapping performance as ZnO-coated MNPs for the same human saliva tryptic digest [42].

2.1.3 *Aluminum Oxide (Al_2O_3)*

Al_2O_3-coated magnetic beads have been described in literature as effective phosphopeptide affinity and sensing tools. Al_2O_3-coated MNPs have trapping capacities of 60 μg of phosphopeptides per milligram of particles, and a detection limit of 25 fmol for human protein phosphatase inhibitor 1 (PPI 1) tryptic digest. The entire process of enrichment and MALDI MS analysis takes approximately 5 min [43]. Al_2O_3-coated MNPs modified with a fluorophore—riboflavin-5′-monophosphate (RFMP-Fe_3O_4@Al_2O_3) have been used as affinity and sensing probes for phosphorylated fibrinopeptide A, a peptide which exists in elevated levels in patients with gastric and ovarian cancers and is known to exist in phosphorylated form to an extent of 20–30 % in human blood. RFMP molecules immobilized at the surface of the particles exchange with phosphorylated fibrinopeptide A in solution, allowing the

quantitative analysis of the solution by fluorescence spectroscopy. Particles with the trapped phosphorylated peptide may subsequently be analyzed qualitatively by mass spectrometry [44].

2.1.4 Other Metal Oxides

Niobium pentoxide (Nb_2O_5) efficiently enriches phosphopeptides from standard tryptic protein mixtures and cellular lysates with 50 % overlap in peptide sequence when compared to TiO_2, which translates into a high degree of orthogonality between both metal oxides. In addition, Nb_2O_5 and TiO_2 presented similar recovery efficiencies of phosphopeptides between 50 and 100 % [45]. Nb_2O_5-coated MNPs are able to trap phosphorylated peptides from tryptic digests of caseins, serum and cell lysate in only 1 min using microwave heating and with a detection limit of 5 fmol [46].

Tantalum pentoxide (Ta_2O_5)-coated MNPs present divergent phosphopeptide trapping selectivity when compared to TiO_2-coated MNPs, with a larger number of unique phosphopeptides being identified by Ta_2O_5-coated MNPs [47]. The addition of 2,5-DHB to the loading solution contributes to the enhanced selectivity towards phosphorylated species, a concept that had been introduced by Larsen and coworkers [25, 47].

Other metal oxides, such as gallium oxide (Ga_2O_3) [48] and tin dioxide (SnO_2) [49], have also been combined with Fe_3O_4 and used in phosphoproteomic experiments.

The sensitivity and enrichment time of different metal oxide-coated MNPs are presented in Table 2. However, it should be taken into consideration that this comparison between different materials is not always straightforward, as experiments are not always performed in the same conditions and in some cases use different instrumentation. One interesting detail to point out is that Fe_3O_4 magnetic particles are excellent microwave absorbers. Microwave heating can be applied both during tryptic digestion and enrichment steps, which in addition to the ease of separation of MNPs makes the entire process extremely time-efficient.

2.2 Rare Earth Metal-Based Magnetic Nanocomposites

Due to the strong affinity between rare earth metal ions and phosphate moieties, a variety of different materials based on this type of metals have been developed for phosphopeptide enrichment. Rare metal ions (hard acids) are able to coordinate oxygen atoms (hard bases) of phosphate groups through mono- or multi-dentate bonds, which make the phosphorous atoms more electropositive and more susceptible to nucleophilic attack by hydroxyl groups. This process results in the cleavage of phosphate-ester bonds and consequent dephosphorylation of the peptides [50].

Very recently, rare earth vanadate-coated MNPs (γ-Fe_2O_3@ $REVO_4$; RE = Sm, Dy, Ho) were used for the first time to enrich phosphorylated peptides from tryptic digests of standard proteins and human serum. The three rare earth metals tested presented similar morphologies, saturation magnetization values, and

Table 2
Sensitivity and enrichment time of different metal oxide-coated magnetic nanocomposites

Nanocomposite	Detection limit	Enrichment time	Instrumentation	Ref.
Fe_3O_4@TiO_2	50 fmol (β-casein)	90 min	Bruker Biflex III (MALDI-TOF)	[33]
Fe_3O_4@C@TiO_2	10 fmol (β-casein)	0.5 min	Applied Biosystems 4700 (MALDI-TOF/TOF)	[34, 35]
γ-Fe_2O_3@TiO_2	10 fmol (β-casein)	30 min	Applied Biosystems Voyager DE-STR (MALDI-TOF)	[36]
Fe_3O_4@ZrO_2	45 fmol (β-casein)	0.5 min	Bruker Biflex III (MALDI-TOF)	[37]
Fe_3O_4@TiO_2-ZrO_2	250 fmol (β-casein)	1 min	Thermo Finnigan LCQ (ESI Ion Trap)	[38]
Yolk-Shell MSP@ZrO_2	2.5 pmol (β-casein)	30 min	Applied Biosystems 5800 (MALDI-TOF)	[40]
Fe_3O_4@ZnO	2.5 fmol (β-casein)	0.5 min	Bruker Biflex III (MALDI-TOF)	[42]
Fe_3O_4@Al_2O_3	25 fmol (α-casein)	0.5 min	Bruker Biflex III (MALDI-TOF)	[43]
Fe_3O_4@Nb_2O_5	5 fmol	1 min	Applied Biosystems 4800 (MALDI-TOF/TOF)	[46]
Fe_3O_4@Ta_2O_5	10 fmol (β-casein)	1 min	Bruker Biflex III (MALDI-TOF)	[74]
Fe_3O_4@C@Ga_2O_3	40 fmol (β-casein)	0.5 min	Applied Biosystems 4700 (MALDI-TOF/TOF)	[48]
Fe_3O_4@C@SnO_2	80 fmol (β-casein)	30 min	Applied Biosystems 4700 (MALDI-TOF/TOF)	[49]

selectivity towards phosphorylated peptides. The sensitivity was slightly worse for γ-Fe_2O_3@$HoVO_4$ (200 fmol for a β-casein tryptic digest) when compared to γ-Fe_2O_3@$SmVO_4$ and γ-Fe_2O_3@$DyVO_4$ (100 fmol). Particles could be used up to five times without significant loss of binding capacity or selectivity [50].

Rare earth metal oxides, such as CeO_2, are also starting to be explored in the field of phosphoproteomics. MNPs coated with an intermediate layer of silica and an outer layer of mesoporous CeO_2 (Fe_3O_4@SiO_2@mCeO_2) are multifunctional probes, as they have phosphate-affinity, magnetic properties, and they catalyze the dephosphorylation of phosphopeptides, which results in a specific neutral loss of $n \times 80$ Da that can be detected in the MS spectra [51]. The same number of phosphorylated peptides and their correspondent label ions (with decreased masses of $n \times 80$ Da) from a tryptic digest of β-casein could be identified using lanthanum silicate coated MNPs (Fe_3O_4@$La_xSi_yO_5$). Moreover, the relative intensity of the multi-phosphorylated peptide at m/z 3122 Da was much higher than for Fe_3O_4@SiO_2@mCeO_2. This is probably

related with the differences in the catalytic efficiency of La and Ce towards phosphate hydrolysis: milder dephosphorylation facilitating phosphopeptide identification [52].

Very recently, a 3D flowerlike structure composed of a γ-Fe_2O_3 magnetic core coated with a shell of ammonium fluoride and lutetium fluoride (γ-Fe_2O_3@xNH_4F.yLuF_3) was used for the selective capture of phosphopeptides from β-casein digest, nonfat milk tryptic digest and human serum. These 3D nanostructured architectures have unique properties such as highly specific surface areas, low density, and large open pores [53].

A yolk-shell nanostructure composed of a Fe_3O_4 magnetic core and an yttrium phosphate (YPO_4) hollow porous affinity shell (Fe_3O_4@hYPO_4) is another example of success in the application of rare metal ions for the enrichment of phosphopeptides, where a detection limit of 10 fmol was determined for β-casein tryptic digests [54]. A distinct approach consisting of coating an ultrathin YPO_4 shell on polyacrylate capped Fe_3O_4 (PA-Fe_3O_4@YPO_4) allowed faster adsorption/desorption dynamics, and low nonspecific binding [55]. Table 3 presents the sensitivity and enrichment times of different rare earth metal-based magnetic nanocomposites.

2.3 IMAC-Based Magnetic Nanocomposites

Despite some inherent disadvantages, such as nonspecific binding and metal leaching, IMAC still plays a prominent role in many purification and enrichment processes. In IMAC, transition metal ions are immobilized onto a solid support using a chelating ligand, such as iminodiacetic acid (IDA), nitrilotriacetic acid (NTA), or tris(carboxymethyl)ethylene diamine (TED). The selection of metal ion depends on the application: divalent cations (Cu^{2+}, Ni^{2+}, Zn^{2+}, Co^{2+}) are used for the purification of histidine-tagged proteins, whereas trivalent ions (Al^{3+}, Ga^{3+}, Fe^{3+}) and the tetravalent ion Zr^{4+} are used for phosphopeptide enrichment [56]. Rare

Table 3
Sensitivity and enrichment time of different rare earth metal-based magnetic nanocomposites

Nanocomposite	Detection limit	Enrichment time	Instrumentation	Ref.
γ-Fe_2O_3@$HoVO_4$	200 fmol (β-casein)	5 min	Bruker Autoflex III (MALDI-TOF)	[50]
γ-Fe_2O_3@$SmVO_4$	100 fmol (β-casein)	5 min	Bruker Autoflex III (MALDI-TOF)	[50]
γ-Fe_2O_3@$DyVO_4$	100 fmol (β-casein)	5 min	Bruker Autoflex III (MALDI-TOF)	[50]
Fe_3O_4@La*x*Si*y*O_5	100 fmol (β-casein)	2 min	Bruker Autoflex III (MALDI-TOF)	[52]
Fe_3O_4@hYPO_4	10 fmol (β-casein)	2 min	AB SCIEX 5800 (MALDI-TOF/TOF)	[54]
PA-Fe_3O_4@YPO_4	8 fmol (β-casein)	20 s	Bruker Autoflex III (MALDI-TOF)	[55]

earth metals, such as La^{3+}, Ho^{3+}, Er^{3+}, and Ce^{4+}, have also been successfully used for the capture of phosphorylated peptides using IMAC [57, 58].

A common method of functionalizing the surface of nanoparticles with IDA groups is by reacting IDA with a silane coupling agent such as 3-glycidoxypropyltrimethoxysilane (GLYMO). GLYMO-IDA can then be grafted onto the surface of silica-coated Fe_3O_4, providing chelate sites for the immobilization of metal ions. The presence of silica prevents iron leaching at acidic pH, and helps dispersing the particles in solution, besides providing functional groups at the surface of the particles [59]. GLYMO-IDA has already been used to attach Fe^{3+} and Ce^{4+} ions at the surface of magnetic silica nanocomposites [58, 60, 59]. Ce^{4+}-magnetic silica nanocomposites provided higher phosphopeptide selectivity than Fe^{3+}-magnetic silica [58].

A different methodology consisted of coating the magnetic iron oxide core with poly(2-hydroxyethylmethacrylate-co-glycidyl methacrylate) (P(HEMA–GMA)), a polymer which endows both hydrophilicity to prevent nonspecific binding of peptides, and oxirane functional groups able to react with diethyl ester of IDA. Fe^{3+} and Ga^{3+} immobilized IDA-modified magnetic nanoparticles successfully enriched the samples in both mono- and multi-phosphorylated peptides with minor interference of acidic peptides [61].

Ga^{3+} and Zr^{4+} immobilized NTA-modified magnetic nanoparticles are able to detect as little as 50 fmol phosphopeptides from tryptic digests of α- and β-caseins, with the entire process taking less than 10 min. Fe_3O_4 particles are first silanized and coated with succinic anhydride, and then functionalized with NTA using carbodiimide chemistry. Fe_3O_4/NTA/Zr^{4+} presents higher phosphopeptide binding capacity than Fe_3O_4/NTA/Ga^{3+} owning to the fact that Zr^{4+} presents higher coordination number and consequently provides a larger number of phosphopeptide binding sites [62].

The urgent need for rapid automated systems that combine pre-concentration and selective extraction of phosphorylated species from complex samples, allowing qualitative and quantitative analysis, led to the appearance of high-throughput platforms. With this purpose, a robotic platform that manipulates magnetic beads in a 96-well format, providing highly selective automated enrichment of phosphopeptides and rapid evaluation of experimental parameters, such as metal/chelator combinations, buffer composition, and sample clean-up conditions, was recently described. A combination of six metal ions (Fe^{3+}, Ga^{3+}, Al^{3+}, Zn^{2+}, Cu^{2+}, ZrO^{2+}) and two chelating agents (IDA and NTA) were screened. Generally speaking NTA outperformed IDA, and best results in terms of number of phosphopeptides identified and selectivity were found for magnetic particles functionalized with NTA and either Fe^{3+} or Ga^{3+}. Experiments were conducted in parallel in 96-well plates and completed in approximately 45 min [63].

Considerable efforts have been made to surpass problems associated with low ligand density and metal-leaching observed for traditional IDA and NTA linkers. Very recently, Ti^{4+}-immobilized multilayer polysaccharide coated magnetic nanoparticles have been reported as an exquisite alternative, presenting an extremely low detection limit of 0.5 fmol, large binding capacities (100 mg/g), an enrichment recovery of 85 %, and rapid magnetic separation (10 s). In addition, these nanocomposites were effective in the enrichment of human serum and nonfat milk. The Fe_3O_4 core is coated with two layers of silica, which is further functionalized first with a thick multilayer polysaccharide consisting of hyaluronate (HA) and chitosan (CS), and second with titanium phosphate (Fe_3O_4@SiO_2@$(HA/CS)_{10}$-Ti^{4+}). The multilayer polysaccharide provides both hydrophilic properties and larger number of immobilized Ti^{4+} ions [64].

Recently, Zhang and coworkers have reported the use of adenosine triphosphate (ATP) as chelating ligand. ATP presents two main advantageous features: (1) it is hydrophilic and therefore minimizes nonspecific binding of non-phosphorylated peptides through hydrophobic interactions; and (2) it is able to immobilize metal ions through intermolecular and intramolecular forces providing cross-linked metal-phosphonate sites for the binding of phosphorylated peptides. Using on-target enrichment it was possible to identify peptides from β-casein tryptic digests at the attomole level [65].

Table 4 presents the sensitivity and enrichment times of different IMAC-based magnetic nanocomposites.

2.4 Polymer-Based Magnetic Nanocomposites

As mentioned in the previous section, coating nanoparticles with polymers may present several advantages, as they tune surface chemistry by introducing different functional groups, help preventing particle agglomeration and reduce nonspecific binding.

Table 4
Sensitivity and enrichment time of different IMAC coated-magnetic nanocomposites

Nanocomposite	Detection limit	Enrichment time	Instrumentation	Ref.
Fe_3O_4@SiO_2@GLYMO-IDA-Fe^{3+}	20 fmol (α-casein)	30 min	Applied Biosystems 4800 (MALDI-TOF/TOF)	[60]
Fe_3O_4@SiO_2@NTA-Zr^{4+} Fe_3O_4@SiO_2@NTA-Ga^{3+}	50 fmol (α- and β-casein)	30 s	Bruker Biflex III (MALDI-TOF)	[62]
Fe_3O_4@SiO_2@$(HA/CS)_{10}$-Ti^{4+}	0.5 fmol (β-casein)	30 min	AB SCIEX 5800 (MALDI-TOF/TOF)	[64]
Fe_3O_4@ATP-Ti^{4+}	3 amol (β-casein)	N.A.	N.A.	[65]

N.A. not available

The choice of polymer is extremely important and several factors must be carefully considered, such as charge, hydrophobicity, molecular weight, conformation, biodegradation, degree of surface coverage [32, 66, 67].

Chen and coworkers have used polyethylenimine (PEI), a branched polymer with a high-density of amine groups and strong protonation capacity, to coat Fe_3O_4 MNPs. PEI-coated MNPs were able to enrich phosphopeptides from tryptic digests of protein mixtures consisting of 0.07 % (mol/mol) phosphoproteins in only 1 min. The sensitivity of the method was determined to be 5 fmol using α- and β-casein tryptic digests. PEI-coated MNPs are positively charged within a wide pH range (3–11) and interact electrostatically with the negatively charged phosphate groups. However, similar to other enrichment methods, factors such as the composition of the binding buffer have visible effects on phosphopeptide binding. When 100 % ACN/0.1 % TFA (v/v) was used as binding solvent, both mono- and multi-phosphorylated peptides were detected on the MS spectra. Increasing the water content of the binding solvent (50 % ACN/0.1 % TFA (v/v)) led to the identification of higher intensity peaks correspondent to multi-phosphorylated peptides, but the mono-phosphorylated species did not appear on spectra. This is speculated to be related to the fact that hydration promotes phosphopeptide binding and, therefore, the preferential binding of multiply phosphorylated peptides. A two-step treatment using both solvents, first 100 % ACN/0.1 % TFA (v/v) and then 50 % ACN/0.1 % TFA (v/v), provided improved results as both mono- and multi-phosphorylated peptides could be identified with higher intensities [68].

A new method coined "polymer-based metal ion affinity capture (PolyMAC)" has been revolutionizing the concept of phosphopeptide enrichment. The method consists of functionalizing a soluble nanopolymer (polyamidoamine (PAMAM) dendrimer generation 4) with phosphonate groups which in turn chelates Ti^{4+} ions. The nanopolymer is also modified with aldehyde ("handle") groups. In total, the PolyMAC-Ti^{4+} reagent contains 35 Ti^{4+} and 6 aldehyde groups, but this number can be tailored by adapting the reaction conditions. PolyMAC-Ti^{4+} captures phosphopeptides in solution phase and is then immobilized onto a solid support, such as agarose or magnetic beads, through the formation of a hydrazone bond between its aldehyde groups and hydrazine groups at the surface of the solid support (*see* Fig. 3). PolyMAC showed better performance in terms of number of unique phosphosites identified and enrichment selectivity when compared to other commercially available TiO_2 and IMAC reagents (*see* Table 1) [69].

2.5 Other Magnetic Nanocomposites

The number of novel magnetic nanocomposites with affinity for phosphorylated peptides is increasing day to day. This section describes some of the recently reported novel composites which showed great potential for phosphopeptide enrichment.

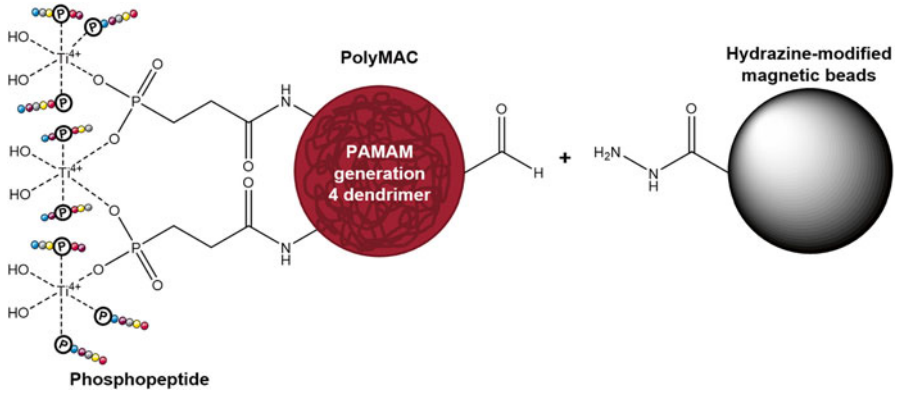


Fig. 3 Phosphopeptide capture using PolyMAC. PolyMAC beads chelate phosphate groups of phosphopeptides through Ti^{4+} metal ions. Phosphopeptide-bound PolyMAC are then reacted with hydrazine-functionalized magnetic beads via aldehyde groups at the surface of the nanopolymer

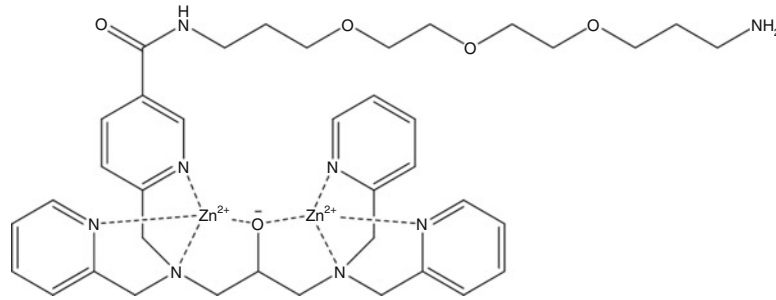


Fig. 4 Illustration of amino-pendant Zn^{2+}-Phos-tag ligand

During the last decade, a few phosphate-affinity ligands based on metal chelation have been reported in the literature [15]. Among them, Phos-tag excelled by its competitiveness to other commercially available traditional implemented IMAC and MOAC based materials. Phos-tag consists of an alkoxide-bridged dinuclear Zn(II) complex with 1,3-bis(pyridin-2-ylmethylamino)propan-2-olate which is able to chelate phosphate moieties through the divalent metal ion (*see* also Chapter 3). Zn^{2+}-Phos-tag was attached to *N*-hydroxysuccinimide-activated agarose-coated magnetic beads using a 15-atom amine-terminated spacer (*see* Fig. 4). One of the advantages of this method is that it allows the use of buffers at physiological pH. Phosphate binding capacity was 4 μmol phenyl phosphate dianion per milliliter Zn^{2+}-Phos-tag magnetic beads. Beads can be used up to 15 times without loss of performance and phosphopeptide recovery yields of nearly 100 % [70].

The chemical stability and excellent mechanical and electronic properties of carbon nanotubes (CNTs) boosted their application in numerous scientific fields [71]. CNTs have been combined with

several solid supports, such as gold and magnetic nanoparticles, because they help preventing macroscopic particle agglomeration. Recently, magnetic CNTs were synthesized via a hydrothermal method and then modified with TiO_2 (MagCNT@TiO_2). TiO_2 was added in order to increase the surface area and due to its phosphate-affinity properties. The composites presented a sensitivity of 20 fmol for a tryptic β-casein digest and could be reused up to ten times [72].

Fe_3O_4 particles derivatized with octadecyltrimethoxysilane (C_{18}-functionalized magnetic beads) have been used to capture both phosphorylated and non-phosphorylated peptides, which can subsequently be selectively desorbed and analyzed by MS without the need of an elution step. C_{18} binds non-phosphorylated peptides through hydrophobic interactions, while the Fe_3O_4 core is able to chelate phosphorylated peptides. Desalting is achieved by washing the particles with 0.1 % formic acid (v/v). After desalting, particles can be resuspended in α-cyano-4-hydroxy-cinnamic acid (α-CHCA) matrix and spotted onto the MALDI target for the identification of non-phosphorylated peptides. The analysis of phosphorylated peptides requires two additional washing steps with 75 % ACN/0.25 % H_2SO_4 and 75 % ACN/1 % NH_4OH in order to remove the non-phosphorylated peptides bound to the particles. The washed particles are then resuspended in 2,5-DHB with 1 % H_3PO_4 and spotted onto the MALDI plate. The choice of an appropriate MALDI matrix is critical for a sensitive and accurate identification of both phosphorylated and non-phosphorylated peptides [73].

3 Conclusions

Magnetic nanocomposites are promising materials in the phosphoproteomics field, because they combine the superparamagnetic properties of the Fe_3O_4/γ-Fe_2O_3 core with the phosphate-affinity properties of different coating materials. In addition, Fe_3O_4/γ-Fe_2O_3 nanoparticles are able to absorb microwave radiation, allowing faster enzymatic digestion and shorter incubation periods. Therefore, these multifunctional composites are optimized for high-throughput protocols, which are both time and cost effective.

Acknowledgments

The authors thank the financial support from Fundação para a Ciência e a Tecnologia, Ministério da Educação e da Ciência, Portugal, through projects no. Pest-OE/EQB/LA0004/2011, PEst-C/EQB/LA0006/2013, PTDC/EBB-BIO/102163/2008, PTDC/EBBBIO/118317/2010 and SFRH/BD/64427/2009 for I.L.B.

References

1. López E, Cho W (2012) Phosphoproteomics and lung cancer research. Int J Mol Sci 13:12287–12314
2. Rigbolt KT, Blagoev B (2012) Quantitative phosphoproteomics to characterize signaling networks. Semin Cell Dev Biol 23:863–871
3. Mann M, Ong S-E, Grønborg M et al (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20:261–268
4. Engholm-Keller K, Larsen MR (2013) Technologies and challenges in large-scale phosphoproteomics. Proteomics 13:910–931
5. Harsha H, Pandey A (2010) Phosphoproteomics in cancer. Mol Oncol 4:482–495
6. Iqbal K, Liu F, Gong C-X et al (2010) Tau in Alzheimer disease and related tauopathies. Curr Alzheimer Res 7:656
7. Braithwaite SP, Stock JB, Mouradian MM (2012) α-Synuclein phosphorylation as a therapeutic target in Parkinson's disease. Rev Neurosci 23:191–198
8. Kotlo K, Johnson KR, Grillon JM et al (2012) Phosphoprotein abundance changes in hypertensive cardiac remodeling. J Proteomics 77:1–13
9. Jaros JA, Martins-de-Souza D, Rahmoune H et al (2012) Protein phosphorylation patterns in serum from schizophrenia patients and healthy controls. J Proteomics 76:43–55
10. Nita-Lazar A, Saito-Benz H, White FM (2008) Quantitative phosphoproteomics by mass spectrometry: past, present, and future. Proteomics 8:4433–4443
11. Eyrich B, Sickmann A, Zahedi RP (2011) Catch me if you can: mass spectrometry-based phosphoproteomics and quantification strategies. Proteomics 11:554–570
12. Palumbo AM, Smith SA, Kalcic CL et al (2011) Tandem mass spectrometry strategies for phosphoproteome analysis. Mass Spectrom Rev 30:600–625
13. Macek B, Mann M, Olsen JV (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol 49:199–221
14. Beltran L, Cutillas PR (2012) Advances in phosphopeptide enrichment techniques for phosphoproteomics. Amino Acids 43: 1009–1024
15. Batalha IL, Lowe CR, Roque AC (2012) Platforms for enrichment of phosphorylated proteins and peptides in proteomics. Trends Biotechnol 30:100–110
16. Villén J, Gygi SP (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3:1630–1638
17. Zarei M, Sprenger A, Metzger F et al (2011) Comparison of ERLIC–TiO_2, HILIC–TiO_2, and SCX–TiO_2 for global phosphoproteomics approaches. J Proteome Res 10:3474–3483
18. Nühse TS, Stensballe A, Jensen ON et al (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Mol Cell Proteomics 2:1234–1243
19. Motoyama A, Xu T, Ruse CI et al (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem 79: 3623–3634
20. Pina AS, Batalha ÍL, Roque ACA (2014) Affinity tags in protein purification and peptide enrichment: an overview. In: Labrou NE (ed) Protein downstream processing, vol 1129, Methods in molecular biology. Humana Press, Totowa, NJ, pp 147–168. doi:10.1007/978-1-62703-977-2_14
21. Cai D, Lee A, Chiang C-M et al (2011) Peptoid ligands that bind selectively to phosphoproteins. Bioorg Med Chem Lett 21:4960–4964
22. Fíla J, Honys D (2012) Enrichment techniques employed in phosphoproteomics. Amino Acids 43:1025–1047
23. Ficarro SB, McCleland ML, Stukenberg PT et al (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. Nat Biotechnol 20:301–305
24. Pinkse MW, Uitto PM, Hilhorst MJ et al (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76:3935–3943
25. Larsen MR, Thingholm TE, Jensen ON et al (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 4:873–886
26. Ye J, Zhang X, Young C et al (2010) Optimized IMAC–IMAC protocol for phosphopeptide recovery from complex biological samples. J Proteome Res 9:3561–3573
27. Najam-ul-Haq M, Jabeen F, Hussain D et al (2012) Versatile nanocomposites in phosphoproteomics: a review. Anal Chim Acta 747: 7–18

28. Zhu Y, Stubbs LP, Ho F et al (2010) Magnetic nanocomposites: a new perspective in catalysis. ChemCatChem 2:365–374
29. Wu W, He Q, Jiang C (2008) Magnetic iron oxide nanoparticles: synthesis and surface functionalization strategies. Nano Res Lett 3:397–415
30. Lee A, Yang HJ, Lim ES et al (2008) Enrichment of phosphopeptides using bare magnetic particles. Rapid Commun Mass Spectrum 22:2561–2564
31. Lu AH, Salabas EL, Schüth F (2007) Magnetic nanoparticles: synthesis, protection, functionalization, and application. Angew Chem Int Ed 46:1222–1244
32. Veiseh O, Gunn JW, Zhang M (2010) Design and fabrication of magnetic nanoparticles for targeted drug delivery and imaging. Adv Drug Deliv Rev 62:284–304
33. Chen C-T, Chen Y-C (2005) Fe_3O_4/TiO_2 core/shell nanoparticles as affinity probes for the analysis of phosphopeptides using TiO_2 surface-assisted laser desorption/ionization mass spectrometry. Anal Chem 77:5912–5919
34. Li Y, Xu X, Qi D et al (2008) Novel Fe_3O_4@ TiO_2 core–shell microspheres for selective enrichment of phosphopeptides in phosphoproteome analysis. J Proteome Res 7:2526–2538
35. Li Y, Wu J, Qi D et al. (2008) Novel approach for the synthesis of Fe_3O_4@TiO_2 core–shell microspheres and their application to the highly specific capture of phosphopeptides for MALDI-TOF MS analysis. Chem Commun 564–566
36. Lu Z, Duan J, He L et al (2010) Mesoporous TiO_2 nanocrystal clusters for selective enrichment of phosphopeptides. Anal Chem 82:7249–7258
37. Lo C-Y, Chen W-Y, Chen C-T et al (2007) Rapid enrichment of phosphopeptides from tryptic digests of proteins using iron oxide nanocomposites of magnetic particles coated with zirconia as the concentrating probes. J Proteome Res 6:887–893
38. Li W, Deng Q, Fang G et al (2013) Facile synthesis of Fe_3O_4@TiO_2–ZrO_2 and its application in phosphopeptide enrichment. J Mater Chem B 1:1947–1961
39. Kweon HK, Håkansson K (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. Anal Chem 78:1743–1749
40. Ma W, Zhang C, Zhang Y et al (2014) Magnetic MSP@ZrO_2 microspheres with yolk-shell structure: designed synthesis and application in highly selective enrichment of phosphopeptides. Langmuir 30:6602
41. Guo J, Yang W, Wang C (2013) Magnetic colloidal supraparticles: design, fabrication and biomedical applications. Adv Mater 25:5196–5214
42. Chen W-Y, Chen Y-C (2010) Functional Fe_3O_4@ZnO magnetic nanoparticle-assisted enrichment and enzymatic digestion of phosphoproteins from saliva. Anal Bioanal Chem 398:2049–2057
43. Chen C-T, Chen W-Y, Tsai P-J et al (2007) Rapid enrichment of phosphopeptides and phosphoproteins from complex samples using magnetic particles coated with alumina as the concentrating probes for MALDI MS analysis. J Proteome Res 6:316–325
44. Chen C-T, Chen Y-C (2010) Functional magnetic nanoparticle-based label free fluorescence detection of phosphorylated species. Chem Commun 46:5674–5676
45. Ficarro SB, Parikh JR, Blank NC et al (2008) Niobium (V) oxide (Nb2O5): application to phosphoproteomics. Anal Chem 80:4606–4613
46. Lin H-Y, Chen W-Y, Chen Y-C (2009) Iron oxide/niobium oxide core–shell magnetic nanoparticle-based phosphopeptide enrichment from biological samples for MALDI MS analysis. J Biomed Nanotechnol 5:215–223
47. Qi D, Lu J, Deng C et al (2009) Development of core-shell structure Fe3O4@ Ta2O5 microspheres for selective enrichment of phosphopeptides for mass spectrometry analysis. J Chromatogr A 1216:5533–5539
48. Li Y, Lin H, Deng C et al (2008) Highly selective and rapid enrichment of phosphorylated peptides using gallium oxide-coated magnetic microspheres for MALDI-TOF-MS and nano-LC-ESI-MS/MS/MS analysis. Proteomics 8:238–249
49. Qi D, Lu J, Deng C et al (2009) Magnetically responsive Fe_3O_4@C@SnO_2 core–shell microspheres: synthesis, characterization and application in phosphoproteomics. J Phys Chem C 113:15854–15861
50. Wang Z-G, Cheng G, Liu Y-L et al (2013) Magnetic γ-Fe_2O_3@$REVO_4$ (RE = Sm, Dy, Ho) affinity microspheres for selective capture, fast separation and easy identification of phosphopeptides. J Mater Chem B 1:1491–1500
51. Cheng G, Zhang J-L, Liu Y-L et al (2011) Synthesis of novel Fe_3O4@SiO_2@CeO_2 microspheres with mesoporous shell for phosphopeptide capturing and labeling. Chem Commun 47:5732–5734
52. Cheng G, Liu Y-L, Zhang J-L et al (2012) Lanthanum silicate coated magnetic microspheres as a promising affinity material for phosphopeptide enrichment and identification. Anal Bioanal Chem 404:763–770

53. Wang Z-G, Cheng G, Liu Y-L et al (2013) Novel 3D flowerlike hierarchical γ-Fe_2O_3@ $xNH_4F \cdot yLuF_3$ core–shell microspheres tailor-made by a phase transformation process for the capture of phosphopeptides. J Mater Chem B 1:4845–4854
54. Cheng G, Liu Y-L, Wang Z-G et al (2013) Yolk–shell magnetic microspheres with mesoporous yttrium phosphate shells for selective capture and identification of phosphopeptides. J Mater Chem B 1:3661–3669
55. Sun Y, Wang H-F (2013) Ultrathin-yttrium phosphate-shelled polyacrylate-ferriferrous oxide magnetic microspheres for rapid and selective enrichment of phosphopeptides. J Chromatogr A 1316:62–68
56. Block H, Maertens B, Spriestersbach A et al (2009) Immobilized-metal affinity chromatography (IMAC): a review. Method Enzymol 463:439–473
57. Mirza MR, Rainer M, Messner CB et al (2013) A new type of metal chelate affinity chromatography using trivalent lanthanide ions for phosphopeptide enrichment. Analyst 138:2995–3004
58. Li Y, Qi D, Deng C et al (2008) Cerium ion-chelated magnetic silica microspheres for enrichment and direct determination of phosphopeptides by matrix-assisted laser desorption ionization mass spectrometry. J Proteome Res 7:1767–1777
59. Xu X, Deng C, Gao M et al (2006) Synthesis of magnetic microspheres with immobilized metal ions for enrichment and direct determination of phosphopeptides by matrix-assisted laser desorption ionization mass spectrometry. Adv Mater 18:3289–3293
60. Tan F, Zhang Y, Mi W et al (2008) Enrichment of phosphopeptides by Fe^{3+}-immobilized magnetic nanoparticles for phosphoproteome analysis of the plasma membrane of mouse liver. J Proteome Res 7:1078–1087
61. Novotna L, Emmerova T, Horak D et al (2010) Iminodiacetic acid-modified magnetic poly (2-hydroxyethyl methacrylate)-based microspheres for phosphopeptide enrichment. J Chromatogr A 1217:8032–8040
62. Li Y-C, Lin Y-S, Tsai P-J et al (2007) Nitrilotriacetic acid-coated magnetic nanoparticles as affinity probes for enrichment of histidine-tagged proteins and phosphorylated peptides. Anal Chem 79:7519–7525
63. Ficarro SB, Adelmant G, Tomar MN et al (2009) Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. Anal Chem 81:4566–4575
64. Xiong Z, Zhang L, Fang C et al (2014) Ti^{4+}-immobilized multilayer polysaccharide coated magnetic nanoparticles for highly selective enrichment of phosphopeptides. J Mater Chem B 2:4473
65. Zhang L, Zhao Q, Liang Z et al (2012) Synthesis of adenosine functionalized metal immobilized magnetic nanoparticles for highly selective and sensitive enrichment of phosphopeptides. Chem Commun 48:6274–6276
66. Sun C, Lee JSH, Zhang M (2008) Magnetic nanoparticles in MR imaging and drug delivery. Adv Drug Deliv Rev 60:1252–1265
67. Dias AMGC, Hussain A, Marcos AS et al (2011) A biotechnological perspective on the application of iron oxide magnetic colloids modified with polysaccharides. Biotechnol Adv 29:142–155
68. Chen C-T, Wang L-Y, Ho Y-P (2011) Use of polyethylenimine-modified magnetic nanoparticles for highly specific enrichment of phosphopeptides for mass spectrometric analysis. Anal Bioanal Chem 399:2795–2806
69. Iliuk AB, Martin VA, Alicie BM et al (2010) In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. Mol Cell Proteomics 9:2162–2172
70. Tsunehiro M, Meki Y, Matsuoka K et al (2013) A Phos-tag-based magnetic-bead method for rapid and selective separation of phosphorylated biomolecules. J Chromatogr B 925:86–94
71. Fang G, Gao W, Deng Q et al (2012) Highly selective capture of phosphopeptides using a nano titanium dioxide–multiwalled carbon nanotube nanocomposite. Anal Biochem 423:210–217
72. Yan Y, Zheng Z, Deng C et al (2014) Selective enrichment of phosphopeptides by titania nanoparticles coated magnetic carbon nanotubes. Talanta 118:14–20
73. Hsiao H-H, Hsieh H-Y, Chou C-C et al (2007) Concerted experimental approach for sequential mapping of peptides and phosphopeptides using C18-functionalized magnetic nanoparticles. J Proteome Res 6:1313–1324
74. Lin H-Y, Chen W-Y, Chen Y-C (2009) Iron oxide/tantalum oxide core–shell magnetic nanoparticle-based microwave-assisted extraction for phosphopeptide enrichment from complex samples for MALDI MS analysis. Anal Bioanal Chem 394:2129–2136

Part IV

Phosphoprotein Analysis by Mass Spectrometry

Chapter 14

Two Dimensional Gel Electrophoresis-Based Plant Phosphoproteomics

Chao Han and Pingfang Yang

Abstract

Phosphorylation is one of the most important reversible protein modifications and is involved in regulating signal transduction, subcellular localization and enzyme activity of target proteins. Phosphorylation or dephosphorylation of proteins is directly reflected in changed ratios of phosphoprotein abundance and total protein abundance. Two-dimensional gel electrophoresis (2-DE)-based proteomics allow quantification of both total protein abundance by Coomassie Brilliant Blue (CBB) staining and phosphoprotein abundance by fluorescence-based staining. Pro-Q diamond phosphoprotein stain (Pro-Q DPS) can bind to the phosphate moiety of the phospho-amino acid directly, regardless of the nature of the phospho-amino acid. Phosphoproteins can thus be detected using proper excitation light, quantified using image analysis software and subsequently be subjected to analysis by mass spectrometry. Here, we describe a protein phosphorylation status analysis method combining both CBB and Pro-Q DPS staining based on 2-DE gel-based phosphoproteomics, which has been widely applied to plant phosphoproteomics studies.

Key words Phosphoproteomics, 2-DE, Pro-Q, Coomassie brilliant blue

1 Introduction

Two dimensional gel electrophoresis (2-DE) combined with mass spectrometry analysis is a classic strategy for proteomics studies, which cannot only be used for protein quantification but also for protein modification analysis. Gel-free proteomics strategies have become more and more popular because of the improvement in mass spectrometry technology. Nevertheless proteomics workflows relying on two dimensions of gel electrophoreses (2-DE), which can separate proteins based on their isoelectric points on the first dimension and molecular weight on the secondary one, are still amongst the most reproducible and robust proteomics setups [1]. 2-DE gel-based proteomics can serve as a platform for protein modification studies by exploiting the fact that many modifications induce a pI shift and thus a separation of the modified protein from the pool of the unmodified one [2]. Pro-Q Diamond dyes can be

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_14,

used for fluorescent detection of phosphorylated proteins directly in sodium dodecyl sulfate (SDS) polyacrylamide gels or 2-DE gels. This fluorescent dye possesses linear signal intensity depending on phosphoprotein abundance and the amount of phosphorylated amino acids in a protein, whose expression does not change [3]. Pro-Q DPS dye, as a commercial product, is a high cost reagent for those methods. Proper dilution and storage conditions as introduced by Ganesh K. Agrawal and Jay J. Thelen [4] are necessary for prolonging the usage period of the dyes. According to their dilution method, we obtained high quality Pro-Q stain gel images for our recent work on rice germinating seeds [5]. Moreover, the staining with Pro-Q dye is compatible with other dyes, such as Coomassie Brilliant Blue (CBB) and SYPRO Ruby, as well as with analysis by mass spectrometry [3, 6]. Therefore, combining 2-DE and Pro-Q staining is an ideal strategy for quantitative phosphoproteomic analysis on a global scale, which has been widely used in large-scale functional protein detection in seed development [7], flooding stress [8], cell dedifferentiation [9] and subcellular organelles [10]. Besides, it is also a tool of detecting phosphorylated isoforms of specific target protein [11]. Here we present a gel-based phosphoproteomic workflow used for plant tissues including 2-DE and Pro-Q staining processes.

2 Materials

Prepare all solutions using ultrapure Milli-Q water (18 M Ω, 25 °C) and analytical grade reagents. All solutions should be stored at proper temperature.

2.1 Protein Extraction: Tris/Acetone Method

1. Mortar and pestle. All mortars and pestles are sterilized at 120 °C, 0.15 MPa for 20 min.
2. Extraction Buffer: 20 mM Tris–HCl (pH 7.5), 250 mM sucrose, 10 mM EGTA, 1 mM protease inhibitor phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT), 1 % Triton X-100, and phosphatase inhibitor 5 mM sodium fluoride, 5 mM sodium orthovanadate, 25 mM glycerophosphate, 10 mM sodium pyrophosphate. Stored at 4 °C. Phosphatase inhibitor, protease inhibitor and DTT are added into Extraction Buffer right before use. PMSF, DTT, sodium fluoride and sodium orthovanadate are prepared as 100 mM, 1 M, 500 mM and 500 mM storage solution, respectively. 1 mL Extraction Buffer is mixed with 10 μL PMSF storage solution, 1 μL DTT storage solution, 10 μL sodium fluoride storage solution, 10 μL sodium orthovanadate storage solution, 0.0054 g glycerophosphate, and 0.0027 g sodium pyrophosphate.

3. Protein Depositing Solution: 100 % Acetone stored at room temperature, precooled at –20 °C.
4. Lysis Buffer: 7 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 65 mM DTT, and Bio-Lyte pH 5–8 (Bio-Rad). Stored at –20 °C.
5. Table top centrifuge.
6. Probe sonicator.
7. Liquid nitrogen.
8. SpeedVac.

2.2 Protein Quantification: Bradford Method

1. Bradford Storage Solution is prepared as follows: mix 100 mL 95 % ethanol, 200 mL 88 % phosphoric acid, 350 mg Coomassie brilliant blue G250 and 200 mL Milli-Q water. Bradford Working Solution is prepared as follows: mix 30 mL Bradford Storage Solution, 30 mL 88 % phosphoric acid, 15 mL 95 % ethanol and 425 mL Milli-Q water.
2. 1 mg/mL bovine serum albumin (BSA) standard solution. Stored at 4 °C.
3. Spectrophotometer with 4 mL glass cell.

2.3 Protein Isoelectric Focusing (IEF)

1. Protein IEF Cell, Protean IEF system with 17 cm disposable trays (e.g., Bio-Rad, Hercules, CA, USA).
2. Linear IPG strips (pH 4–7, 17 cm, Bio-Rad).
3. Mineral oil.
4. 1 % (w/v) bromophenol blue.
5. Filter paper.

2.4 SDS-PAGE

1. Vertical electrophoresis Protean II xi Cell (Bio-Rad).
2. Acrylamide stock solution: 30 % (w/v) acrylamide, 0.8 % *N,N′*-methylene bis-acrylamide. Stored at 4 °C.
3. 1.5 M Tris–HCl, pH 8.8, stored at room temperature.
4. 10 % (w/v) SDS stored at room temperature.
5. TEMED (*N,N,N′,N′*-tetramethyl-ethylenediamine) stored at 4 °C, use fresh.
6. 10 % (w/v) ammonium persulfate stored at –20 °C.
7. SDS Equilibration Buffer: 37.5 mM Tris–HCl pH 8.8, 6 M (w/v) urea, 20 % (v/v) glycerol, 2 % (w/v) SDS. Stored in single-use aliquots at –20 °C in 10 mL Falcon tubes.
8. Reduction Solution: 2 % (w/v) DTT in SDS Equilibration Buffer. Prepare fresh before use.
9. Alkylation Solution: 2.5 % (w/v) iodoacetamide in SDS Equilibration Buffer. Prepare freshly before use.

10. Agarose Sealing Solution: 0.5 % (w/v) low melting point agarose, 25 mM Tris–HCl (pH 7.5), 192 mM glycine, 0.1 % (w/v) SDS, 0.001 % (w/v) bromophenol blue. Stored at room temperature.
11. Electrophoresis Buffer: 25 mM Tris–HCl (pH 7.5), 192 mM glycine, 0.1 % SDS. Stored at room temperature.

2.5 Phosphoprotein Staining Using Pro-Q Dye

1. Plastic gel staining box, size 20 cm × 20 cm × 5 cm.
2. Shaker.
3. Pro-Q DPS dye (Invitrogen, Carlsbad, CA, US). Stored at 4 °C (*see* **Note 1**).
4. Fixation Solution: 50 % (v/v) methanol, 10 % (v/v) acetic acid. Stored at room temperature.
5. Staining Solution: 3× diluted Pro-Q DPS dye (v/v) in Milli-Q water. Prepare freshly before use.
6. Destaining Solution: 50 mM sodium acetate-acetic acid pH 4.0, 20 % (v/v) acetonitrile. Prepare stock solution (1 M sodium acetate-acetic acid pH 4.0). To prepare 1 L of Destaining Solution, combine 50 mL of stock solution, 750 mL Milli-Q water and 200 mL of acetonitrile.

2.6 Protein Staining Using Coomassie Brilliant Blue

1. CBB Staining Solution: 0.116 % Coomassie Brilliant Blue (CBB) R250, 25 % ethanol, 8 % acetic acid. Stored at room temperature.
2. CBB Destaining Solution: 25 % ethanol, 8 % acetic acid. Stored at room temperature.

2.7 2-DE Gel Image Obtaining

1. Laser scanner Typhoon 5600 scanner.
2. Image scanner.

2.8 Gel Image Analysis

1. PDQuest software.

2.9 Protein Spots Excision and In-Gel Digestion

1. Speed Vac Concentrator system (e.g., Labconco, Kansas, USA).
2. Discolor Solution: 25 mM ammonium acid carbonate (NH_4HCO_3), 50 % acetonitrile.
3. Desiccation Solution: 50 % acetonitrile.
4. Trypsin Storage Solution: 20 μg sequence grade trypsin diluted in 50 mM acetic acid. Stored at −80 °C.
5. Trypsin Working Solution: Trypsin Storage Solution diluted 10× with 25 mM NH_4HCO_3. Freshly prepared before use.
6. Peptide Extraction Solution I: 0.1 % trifluoroacetic acid in acetonitrile. Freshly prepared before use.
7. Peptide Extraction Solution II: 0.1 % trifluoroacetic acid. Freshly prepared before use.

8. Peptide Extraction Solution III: 0.1 % trifluoroacetic acid, 50 % acetonitrile. Freshly prepared before use. Prepare the Peptide Extraction Primary Buffer A (1 % trifluoroacetic acid in acetonitrile) and B (1 % trifluoroacetic acid). To get 1 mL Peptide Extraction Buffer I, mix 100 μL Peptide Extraction Primary Buffer A and 900 μL acetonitrile. For preparing 1 mL Peptide Extraction Buffer II, mix 100 μL Peptide Extraction Primary Buffer B and 900 μL Milli-Q water. To get 1 mL Peptide Extraction Buffer III, mix 100 μL Peptide Extraction Primary Buffer B, 400 μL Milli-Q water, and 500 μL acetonitrile.

2.10 Mass Spectrometry Analysis

1. Peptide Solving Solution: 0.1 % trifluoroacetic acid, 50 % acetonitrile.
2. 10 mg/mL α-Cyano-4-hydroxycinnamic acid.
3. Target plate.
4. MALDI-TOF/TOF analyzer (e.g., AB SCIEX, Foster City, USA; Type 5800).
5. Data analysis software.

3 Methods

A proper protein extraction method for experimental material and the gel size are crucial factors for perfect protein separation in the gel system. Proper protein separation is beneficial for phosphoprotein staining, which has a positive effect on phosphoprotein abundance quantification in the Pro-Q-stained gel image. The Tris/Acetone protein extraction method introduced here is suitable for various plant tissues, including rice seed, soybean seed, soybean pistil, maize leaf and endosperm, rape stem and root [12–15]. Larger-sized gels, ranging from 17 cm to 24 cm, are favorable for separation of proteins which possess close isoelectric points and molecular weights [3, 7]. Both biological and experimental replicates are required for downstream data analysis, in order to confirm that phosphoprotein spots emerge at constant positions in the 2-DE gel.

3.1 Protein Extraction from Plant Tissues with Tris/Acetone Method

1. Mince plant tissues (0.1–0.2 g) in liquid nitrogen and transfer the powder to a 50 mL falcon tube containing 5 mL of Extraction Buffer (*see* **Note 2**).
2. Sonicate for 10 min.
3. Centrifuge at 12,000 × *g* at 4 °C for 20 min and transfer the supernatant to a new falcon tube.
4. Add 15 mL Protein Depositing Solution and keep at −20 °C for 2 h.
5. Centrifuge at 12,000 × *g* at 4 °C for 20 min and remove supernatant.

6. Add 1 mL Protein Depositing Solution and resuspend the pellet thoroughly.
7. Transfer the solution together with the pellet to a fresh 2 mL tube.
8. Wash twice with Protein Depositing Solution.
9. Discard the supernatant and dry the pellet by evaporation with CentriVap Concentrator system at 4 °C for 5 min.
10. Resuspend the dried protein powder in 0.5 mL Lysis Buffer and store at −20 °C (*see* **Note 3**).

3.2 Protein Quantification: Bradford Method

1. Prepare the protein standard using a BSA Standard Solution. Add 2.5, 5, 7.5, 10, 12.5, 15, 17.5, and 20 μL 1 mg/mL BSA standard solution into 4 mL tubes.
2. Add the proper volume of Lysis Buffer into each tube to a final volume of 100 μL.
3. Add 100 μL Lysis Buffer into a new tube as the blank sample.
4. Add 2.9 mL Bradford Working Solution into each tube and keep at room temperature for 5 min.
5. After blanking on spectrophotometer, measure the standard sample absorbance at 595 nm and produce a standard curve for protein concentration calculation.
6. Add 5 μL protein sample in Lysis Buffer and 95 μL Lysis Buffer into a 4-mL tube.
7. Add 2.9 mL Bradford Working Solution to the protein sample and keep at room temperature for 5 min.
8. Measure standard sample absorbance at 595 nm and calculate the sample concentration (*see* **Note 4**).

3.3 Protein Isoelectric Focusing

1. Add 1 mg of protein into a 1.5-mL microfuge tube and bring the final volume to 330 μL with Lysis Buffer.
2. Add 0.5 μL 1 % (w/v) bromophenol blue and vortex briefly.
3. Centrifuge at 12,000 × *g* for 5 min.
4. Pipette the supernatant into the IPG rehydrating tray from the left side to the right side (*see* **Note 5**).
5. Peel apart the dehydrated IPG strip (pH 4–7; 17 cm) and place the dried acrylamide side facing downward onto the sample solution in the IPG focusing tray.
6. Keep at room temperature for 1–2 h.
7. After all the sample solution is absorbed into the strip, add 2 mL of mineral oil into the channel of the rehydration tray to cover the strip. Keep at room temperature for 16 h (*see* **Note 6**).
8. After rehydrating, place the strip into the IPG focusing tray, which has already been set with wet filter paper pieces onto the wire electrode.

9. Add 2 mL of mineral oil into the channel of the focusing tray. Finally put the IPG focusing tray into the Protean IEF system.
10. Use a four-step focusing protocol: 50–200 V for 1 h; 200–500 V for 1.5 h; 500–8000 V for 5 h; 8000 V for a total of 42,000 Vh.

3.4 SDS-PAGE

1. Clean the gel polymerization glass plate using deionized water and set the plate well for gel polymerization (*see* **Note 7**).
2. For each 17 cm gel, add 16 mL acrylamide stock solution, 13.2 mL Milli-Q water, 10 mL 1.5 M Tris–HCl pH 8.8, 0.4 mL 10 % SDS, 0.4 mL 10 % ammonium persulfate and 32 μL TEMED. After mixing with a stirring rod for several seconds, pour the gel solution into the glass plates and leave 0.5 cm empty space on the top boundary of the glass plates (*see* **Note 8**).
3. Pipette 1–2 mL Milli-Q water gently on top of the gel solution. Keep at room temperature for polymerization until there is a clear boundary between gel and water. Usually, this takes about 40 min (*see* **Note 9**).
4. After focusing remove the strip from IPG focusing tray.
5. Wash the strip in Milli-Q water for a few seconds in order to remove mineral oil.
6. Put the strip into a clean rehydrating tray and with the gel side facing up.
7. Add 5 mL Reduction Solution for each strip and shake gently at room temperature for 17 min.
8. Put the strip into a new channel of the rehydrating tray, add 5 mL Alkylation Solution and shake gently at room temperature for 15 min.
9. After washing the strip in Milli-Q water for a few seconds, put the strip on the top of the polymerized acrylamide gel (*see* **Note 10**).
10. Add 1 mL heated and melted Agarose Sealing Solution to cover the strip.
11. After the agarose solidified, set the glass plate containing the gel and the strip onto a vertical electrophoresis Protean II xi Cell.
12. Add 1.5–2 L Electrophoresis Buffer into electrophoresis cell.
13. Follow a two step electrophoresis protocol: 100 V for 20 min; 200 V for 6 h.

3.5 Phosphoprotein Staining Using Pro-Q Dye

1. Take out the glass mold from the electrophoresis cell.
2. Transfer the gel to the staining box carefully.
3. Wash the gel twice with 250 mL of Milli-Q water for 5 min each.

4. Incubate at room temperature on a shaker at a speed of 60 rpm for all steps of the Pro-Q staining process.
5. Put the gel into 200 mL of Fixation Solution and incubate for at least 1 h.
6. Wash the gel twice with 250 mL of Milli-Q water for 10 min each.
7. Incubate the gel in the dark for 2 h in 150 mL of Staining Solution containing the Pro-Q dye.
8. Put the gel in 250 mL of Destaining Solution and incubate in the dark for 30 min.
9. Repeat this step three more times for at least 2 h each.
10. Wash the gel with 250 mL of Milli-Q water in the dark for 5 min. Repeat one more time.
11. Scan the gel using laser scanner Typhoon 5600 with 532 nm excitation and 580 nm emission filters. Collect the data at 100 μm resolution and export TIFF files.

3.6 Protein Staining Using Coomassie Brilliant Blue

1. After scanning, immerse the gel in CBB R-250 Staining Solution on a shaker at a speed of 40 rpm overnight at room temperature.
2. Transfer the gel into CBB R-250 Destaining Solution and keep shaking at a speed of 40 rpm at room temperature.
3. Change the CBB R-250 Destaining Solution frequently until the background color on the gel is faded.
4. Scan and analyze the gel using a scanner. Collect data at 600 dpi resolution and export TIFF files.

3.7 Protein Spot Excision and In-Gel Digestion

1. Gel images of Pro-Q staining and CBB staining are analyzed using PDQuest software. Student *t-test* analysis is used to determine significantly changed phosphoprotein spots.
2. The gel spots containing phosphoproteins are excised with a pipette tip and kept in 1.5-mL microfuge tubes.
3. After washing with Milli-Q water, incubate gel spots in 50 μL Discolor Solution for 20 min at 37 °C. Shake the tube intermittently during this process until the color of the spot has faded.
4. Remove the Discolor Solution. Add 50 μL of Desiccation Solution and incubate for 5 min at room temperature.
5. Remove the Desiccation Solution and dry the spot in a CentriVap Concentrator system at 4 °C for 5 min.
6. Add 5 μL Trypsin Working Solution and incubate at 4 °C for 1 h until gel spot is rehydrated.

7. Add 30 μL 25 mM NH_4HCO_3 and incubate at 37 °C for 16 h.
8. After digestion, collect the digested solution (supernatant) into a new 1.5 mL microfuge tube.
9. Add the Peptide Extraction Buffer I to the gel spot tube to extract the residual peptide from the gel spot.
10. Sonicate 10 min at room temperature and transfer the extraction solution into the 1.5-mL microfuge tube from **step 8**.
11. Repeat the last step using Extraction Buffer II and III, respectively.
12. Finally, all extraction solutions are collected into the 1.5-mL microfuge tube from **step 8**.
13. The collected peptide solution is dried in a SpeedVac system at 4 °C for 2 h.

3.8 Mass Spectrometry Analysis and Data Analysis

1. Dissolve the tryptic peptides in 5 μL of 0.1 % trifluoroacetic acid, 50 % acetonitrile.
2. Mix with 5 μL 10 mg/mL α-Cyano-4-hydroxycinnamic acid in 0.1 % trifluoroacetic acid, 50 % acetonitrile.
3. Spot peptides on a target plate.
4. Analyze on a 5800 MALDI-TOF/TOF analyzer (ABI). After calibration, parent mass peaks are scanned using 1000 laser shots with a mass range of 800–4000 Da. Ten parent mass peaks with the highest intensity are picked for tandem TOF/TOF analysis, each with 1500 laser shots.
5. Search MS/MS data against suitable database using Mascot software.
6. Assign search parameters for the database and then define protein assignment criteria to identify phosphoproteins. Search parameters are set as follows: fixed modifications of carbamidomethylation on C and variable modifications of oxidation on M, up to one missed cleavage, precursor ion tolerance at 100 p.p.m., MS/MS fragment ion tolerance at 0.4 Da and peptide charge of 1+. The matched proteins with scores >70, and protein CI% >95 % are selected as identified proteins.

4 Notes

1. Storage of Pro-Q DPS dye at 4 °C will prolong its stability around twofold. Threefold dilution of dye is still sufficient for obtaining high quality gel images.
2. In order to extract sufficient protein from plant tissue, the ratio of Extraction Buffer volume and plant tissue weight should be more than 10:1 (mL/g).

3. After measuring primary protein concentration, dilute the sample to around 3.5–4 mg/mL based on primary concentration and then measure the sample concentration again and record. The result is used for sample loading.
4. The resuspended proteins can be stored at −20 °C for 1 week. For longer storage, protein samples should be transferred to a −80 °C freezer.
5. Pipetting the supernatant into the IPG rehydrating tray will generate bubbles, which can impair loading of the sample onto the strip. Pipette the solution on one side of tray continuously and gently to avoid creating bubbles. Make sure that the protein sample solution forms an even liquid layer on one side of tray. This liquid layer is localized in the middle of the tray, which leaves sufficient empty space at the left and right ends of IPG rehydrating tray. Break bubbles using tweezers.
6. When placing the IPG strip onto the protein sample solution, put one side of the IPG strip on the empty space of the rehydrating tray, left end first. Release the strip gradually and make the strip cover the protein sample solution from left to right. This will avoid generating bubbles between strip and protein sample solution. In order to avoid shifting the strip when adding mineral oil into the tray, pipette the oil into the left end of the tray. The oil will flow to the other end of the tray automatically and gently.
7. Make sure to clean the glass plates thoroughly with deionized water, which avoids generating tiny crests on the surface of the gel during polymerization.
8. Acrylamide stock solution is a nervous toxin, make sure not to spill any solution on skin or clothes and wear personal protective equipment.
9. During the gel polymerization, make sure that there is sufficient water on the top of the gel to avoid the gel from drying. A stacking gel can be added if necessary.
10. Paste the plastic side of the IPG strip on the inside of the glass plate and push the strip on top of the SDS-PAGE gel. Use a hard paper to press the gel gently in order to remove the bubble between strip and gel.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (NSFC, No. 31271805).

References

1. Hackett M (2008) Science, marketing and wishful thinking in quantitative proteomics. Proteomics 8:4618–4623
2. Rabilloud T (2014) How to use 2D gel electrophoresis in plant proteomics in plant proteomics. In: J.V Jorrin-Novo et al. (ed) Plant Proteomics: Methods and Protocols, 2nd edn. Humana Press, NY, p. 43–50
3. Agrawal GK, Thelen JJ (2005) Development of a simplified, economical polyacrylamide gel staining protocol for phosphoproteins. Proteomics 5:4684–4688
4. Agrawal GK, Thelen JJ (2009) A high-resolution two dimensional gel-and Pro-Q DPS-based proteomics workflow for phosphoprotein identification and quantitative profiling. In: M.D Graauw (ed) Phospho-Proteomics. Humana Press, NY, p. 3–19
5. Han C, Wang K, Yang P (2014) Gel-based comparative phosphoproteomic analysis on rice embryo during germination. Plant Cell Physiol 2014, pcu060
6. Steinberg TH, Agnew BJ, Gee KR et al (2003) Global quantitative phosphoprotein analysis using multiplexed proteomics technology. Proteomics 3:1128–1144
7. Agrawal GK, Thelen JJ (2006) Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. Mol Cell Proteomics 5:2044–2059
8. Nanjo Y, Skultety L, Uvácková L et al (2011) Mass spectrometry-based analysis of proteomic changes in the root tips of flooded soybean seedlings. J Proteome Res 11:372–385
9. Chitteti BR, Peng Z (2007) Proteome and phosphoproteome dynamic change during cell dedifferentiation in Arabidopsis. Proteomics 7:1473–1500
10. Ito J, Taylor NL, Castleden I et al (2009) A survey of the Arabidopsis thaliana mitochondrial phosphoproteome. Proteomics 9: 4229–4240
11. Deng Z, Zhang X, Tang W et al (2007) A proteomics study of brassinosteroid response in Arabidopsis. Mol Cell Proteomics 6: 2058–2071
12. He D, Han C, Yao J et al (2011) Constructing the metabolic and regulatory pathways in germinating rice seeds through proteomic approach. Proteomics 11:2693–2713
13. Han C, Yin X, He D et al (2013) Analysis of proteome profile in germinating soybean seed, and its comparison with rice showing the styles of reserves mobilization in different crops. PLoS One 8, e56947
14. Li M, Sha A, Zhou X et al (2012) Comparative proteomic analyses reveal the changes of metabolic features in soybean (Glycine max) pistils upon pollination. Sex Plant Reprod 25: 281–291
15. Méchin V, Damerval, C, Zivy M (2007) Total protein extraction with TCA-acetone. In: H Thiellement et al. (ed) Plant Proteomics. Humana Press, Totowa, NJ, p. 1–8

Chapter 15

Variable Digestion Strategies for Phosphoproteomics Analysis

Humberto Gonczarowska-Jorge, Margherita Dell'Aica, Clarissa Dickhut, and René P. Zahedi

Abstract

In recent years, mass spectrometry-based phosphoproteomics has propelled our knowledge about the regulation of cellular pathways. Nevertheless, typically applied bottom-up strategies have several limitations. Trypsin, the preferentially used proteolytic enzyme shows impaired cleavage efficiency in the vicinity of phosphorylation sites. Moreover, depending on the frequency and distribution of tryptic cleavage sites (Arg/Lys), generated peptides can be either too short or too long for confident identification using standard LC-MS approaches. To overcome these limitations, we introduce an alternative and simple approach based on the usage of the nonspecific serine protease subtilisin, which enables a fast and reproducible digestion and provides access to "hidden" areas of the proteome. Thus, in a single LC-MS experiment >1800 phosphopeptides were confidently identified and localized from 125 μg of HeLa digest, compared to >2100 sites after tryptic digestion. While the overlap was less than 20 %, subtilisin allowed the identification of many phosphorylation sites that are theoretically not accessible via tryptic digestion, thus considerably increasing the coverage of the phosphoproteome.

Key words Subtilisin, Phosphopeptide enrichment, Titanium dioxide, Trypsin, Phosphorylation

1 Introduction

Protein phosphorylation is a reversible posttranslational modification inducing conformational changes in protein structure, which in turn can alter protein activity and function. Thus, it plays a central role in the dynamic regulation of cellular processes and acts as a molecular switch in many different pathways. It is estimated that one third of all proteins in eukaryotic cells are phosphorylated during their life span, and although the analysis of protein phosphorylation by mass spectrometry (usually referred to as phosphoproteomics) has boosted our knowledge about cellular pathways in health and disease, the field still has to face certain limitations [1–4]. Some of these limitations derive from the typical bottom-up proteomic workflow that usually comprises proteolytic digestion with

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_15,

trypsin, followed by enrichment of phosphorylated peptides and subsequent LC-MS analysis. Although in proteomics trypsin is considered as the protease-of-choice, it has certain limitations with regard to specificity and efficiency [5–11]. In addition, protein phosphorylation in proximity to the proteolytic cleavage site can dramatically impair cleavage efficiency, particularly rendering the assessment of phosphorylation stoichiometry extremely challenging—even if stable isotope labeled reference peptides are used for quantitation [12, 13]. When aiming at analyzing the "complete" phosphoproteome of a given sample, the usage of a single protease implies certain limitations. Thus, certain parts of the proteome cannot be covered with trypsin as it generates peptides that are either too short (<7 amino acids) or too long (>35 amino acids) for routine LC-MS methods, impeding peptide identification and phosphorylation site localization, respectively [14].

To overcome the limitations of tryptic digestion, we introduce a simple methodology based on the usage of the serine protease subtilisin [15]. Under optimized conditions subtilisin can be used for reproducible digestion and furthermore provides access to new phosphorylation sites that are concealed from LC-MS identification after tryptic digestion. Consequently, subtilisin provides access to novel phosphorylation sites and thus considerably increases the coverage of the phosphoproteome.

2 Materials

All solutions should be prepared using ultrapure deionized water and stored at 4 °C, if not indicated otherwise. Use LC-MS or UPLC grade for HPLC solvents.

2.1 Cell Culture and Lysis

1. Cell medium: Dulbecco's Modified Eagle's Medium (DMEM) with L-glutamine, 10 % fetal bovine serum, 1 % (v/v) penicillin–streptomycin.
2. Cell line: HeLa S3 cells (DSMZ, Braunschweig, Germany), grown under standard conditions in 75 cm^2 cell culture flask (*see* **Note 1**).
3. Phosphate saline buffer (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4.
4. Cell Detachment Solution: 0.05 % Trypsin, 0.02 % ethylenediaminetetraacetic acid (EDTA) in PBS.
5. Cell Lysis Buffer: 50 mM Tris–HCl, 150 mM NaCl, 1 % (w/v) sodium dodecyl sulfate (SDS), pH 7.8 (adjust with HCl), protease inhibitor cocktail complete Mini EDTA free, phosphatase inhibitor cocktail phosSTOP (*see* **Note 2**).
6. Benzonase >99 % purity for hydrolysis of nucleic acids.
7. Thermomixer.

8. Determination of protein concentration: BCA (*bicinchoninic* acid) protein assay.
9. 15 mL falcon tubes.
10. Thermomixer.
11. Tabletop centrifuge.
12. LoBind Eppendorf tube.
13. SpeedVac.
14. Vortex mixer.

2.2 Carbamidomethylation (CMC) and Ethanol (EtOH) Protein Precipitation

1. Reduction Buffer: stock solution of 2 M dithiothreitol (DTT) in water, can be stored at −40 °C.
2. Alkylation Buffer: stock solution of 1 M iodoacetamide (IAA) in water (*see* **Note 3**).
3. EtOH precipitation: 100 % cold EtOH, store at −40 °C before use.

2.3 Enzymatic Digestion with Subtilisin

1. Subtilisin Solution: dissolve subtilisin P5380 (Sigma-Aldrich, Seelze, Germany) in 50 mM ammonium bicarbonate (ABC), pH 7.8, to a final concentration of 1 μg/μL.
2. Solubilization Buffer: 2 M guanidine hydrochloride (GuHCl) in 50 mM ABC, pH 7.8 (adjust pH with HCl).
3. Digestion Buffer: dilute the solubilization buffer to a final concentration of 0.2 M GuHCl with 50 mM ABC, pH 7.8 and add the required volume of subtilisin solution to obtain a subtilisin–protein ratio of 1:20 (w/w).
4. Stop Solution: 10 % (v/v) trifluoroacetic acid (TFA) (*see* **Note 4**).

2.4 Enzymatic Digestion with Trypsin

1. Trypsin Solution: Sigma trypsin T-1426 (Sigma-Aldrich), lyophilized. Dissolve trypsin in 50 mM ammonium bicarbonate (ABC), pH 7.8, to a final concentration of 1 μg/μL.
2. 1 M $CaCl_2$ Stock Solution.
3. Solubilization Buffer, Digestion Buffer (here with trypsin) and Stop Solution are as above.

2.5 Digestion Control by Monolithic Reversed Phase Chromatography

1. HPLC: UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Germering, Germany) or similar nano HPLC system.
2. Monolithic columns: monolithic trap and main column (used: PepSwift monolithic trap column, 200 μm × 5 mm and PepSwift monolithic capillary column, 200 μm × 5 cm (Thermo Fisher Scientific)).
3. Solvent A: 0.1 % TFA.
4. Solvent B: 0.08 % TFA, 84 % acetonitrile (ACN).

2.6 Titanium Dioxide (TiO_2) Phosphopeptide Enrichment

1. Adsorption material: titanium dioxide beads, 5 μm.
2. Loading Buffer 1: 80 % ACN, 5 % TFA, and 1 M glycolic acid (*see* **Note 5**).
3. Washing Buffer 1: 80 % ACN, 1 % TFA.
4. Washing Buffer 2: 10 % ACN, 0.1 % TFA.
5. Elution Buffer: 1 % (v/v) ammonium hydroxide, pH 11.3.
6. Loading Buffer 2: 70 % ACN, 2 % TFA.
7. Washing Buffer 3: 50 % ACN, 0.1 % TFA.
8. Acidification: 100 % formic acid (FA) and 10 % TFA.

2.7 Sample Purification by Solid Phase Extraction

1. Solid phase: C18 Empore material (3 M, Neuss, Germany) and Oligo R3 material (Applied Biosystems, Foster City, USA).
2. Reversed phase (RP) Washing Buffer: 0.1 % TFA.
3. RP Elution Buffer: 70 % ACN, 0.1 % TFA.

2.8 Reversed Phase Chromatographic Separation and Mass Spectrometry

1. HPLC: UltiMate 3000 nano RSLC system (Thermo Fisher Scientific) or similar nano HPLC system.
2. RP columns: C18 trap and main column (used: Acclaim C18 PepMap100 nano viper trap column, 100 μm × 2 cm, Acclaim C18 PepMap100 nano viper main column, 75 μm × 50 cm (Thermo Fisher Scientific).
3. HPLC Loading Buffer: 0.1 % TFA.
4. HPLC Solvent A: 0.1 % FA.
5. HPLC Solvent B: 0.1 % FA, 84 % ACN.
6. Mass spectrometer: Q Exactive Plus mass spectrometer (Thermo Scientific, Bremen, Germany), or another mass spectrometer that can provide high mass accuracy MS/MS data.

2.9 Data Analysis

1. Data analysis software: Proteome Discoverer (version 1.3, Thermo Scientific).
2. Search algorithm: Mascot algorithm (version 2.4.1, Matrix Science).
3. Phosphorylation site assignment algorithm: phosphoRS 2.0 [16].
4. False discovery rate (FDR) estimation: Peptide Validator.
5. Software for in silico tryptic digestion: DBtoolkit (version 4.2.3, http://code.google.com/p/dbtoolkit/) [17].

3 Methods

3.1 Cell Culture and Lysis

1. Grow HeLa cells in 75 cm^2 cell culture flask in DMEM at 37 °C in a humidified atmosphere with 5 % CO_2 to a confluence of 70–80 %.

2. Remove medium and wash cells once with 5 ml of PBS buffer. Detach cells using 2 ml of trypsin/EDTA Solution and incubate at 37 °C for 3–5 min. Add 8 ml of fresh medium to inactivate trypsin and transfer cell suspension to a falcon tube. Pellet cells by centrifugation (max. 300 × *g*) and wash cell pellet one to two times with PBS.
3. Lyse cells by addition of Lysis Buffer (approximately 500 µL Lysis Buffer per 2 mg of cells).
4. Add benzonase (1.5 µL/100 µL of Lysis Buffer). Add $MgCl_2$ Stock Solution to a final concentration of 2 mM. Incubate for 30 min at 37 °C under gentle shaking in a Thermomixer. If the sample is not clear or still viscous, continue for another 10 min.
5. Centrifuge the sample for 30 min at 4 ° C, at 18,000 × *g*. Transfer the supernatant preferentially to a LoBind Eppendorf tube and discard the pellet.
6. Proceed with a BCA assay to determine the protein concentration using the manufacturer's protocol. This step is crucial for the reproducibility of the experiment. Therefore, for more reliable results, measure at least three serial dilutions of each sample and prepare everything in triplicate (*see* **Note 6**).

3.2 Sample Preparation: Carbamidomethylation, Protein Purification and Digestion

For evaluating the reproducibility of the workflow, perform every digestion in triplicate.

1. Add DTT to a final concentration of 10 mM and incubate at 56 °C for 30 min to reduce cysteines.
2. To prevent disulfide bond rearrangement, alkylate free sulfhydryl groups by the addition of IAA to a final concentration of 30 mM. Incubate for 20 min at room temperature in the dark. Always prepare IAA freshly.
3. For removal of SDS, precipitate proteins by adding 9 volumes of cold ethanol (–40 °C) to 1 volume of sample in lysis buffer. Incubate for 60 min at –40 °C and centrifuge for 30 min at 4 °C and 18,000 × *g*. Carefully discard the supernatant.
4. Resolubilize the protein pellet in Solubilization Buffer. Use approximately 200 µL of Solubilization Buffer for 2 mg of protein. Pipette up and down until the pellet is completely solubilized. If necessary, add more Solubilization Buffer.
5. Dilute the sample to a final concentration of 0.2 M GuHCl using 50 mM ABC buffer. For each sample, save a pre-digestion aliquot corresponding to 1 µg of protein for subsequent digestion control using monolithic HPLC (*see* **Note 7**).
6. Add proteolytic enzyme solution to a final enzyme–protein ratio of 1:20 (w/w). In case of subtilisin, immediately incubate at 56 °C for 20 min. In case of trypsin, add $CaCl_2$ to a final concentration of 2 mM and then incubate at 37 °C for 12 h.

7. After incubation, add 10 % TFA to a final concentration of 1 % in order to stop the digestion.
8. For each sample, save a post-digestion aliquot corresponding to 1 μg of digested protein for subsequent digestion control.
9. Perform a digestion control using the monolithic RP separation on an Ultimate 3000 HPLC. Therefore, for each sample, take the 1 μg aliquots before and after digestion and dilute each sample to 15 μL (loading volume for the HPLC autosampler) with 0.1 % TFA (Solvent A) [5] (*see* **Note 8**).

3.3 TiO_2 Phosphopeptide Enrichment

This protocol is based on a TiO_2 phosphopeptide enrichment protocol published by Larsen and coworkers [18], (*see* also Chapter 9).

1. After digestion, dry samples completely under vacuum.
2. Redissolve the sample in 1 mL of freshly prepared Loading Buffer 1.
3. For each 100 μg of sample, weigh in 1050 μg of TiO_2 beads. Resuspend the beads in a defined volume of Loading Buffer 1, e.g., add 52.5 μL of Loading Buffer 1–1050 μg of TiO_2 beads, to achieve a final concentration of 20 μg of beads/μL of buffer.
4. For a first round of enrichment, add TiO_2 beads in a bead–peptide ratio of 6:1, e.g., 600 μg (corresponding to 30 μL) of beads to 100 μg of peptides, to the dissolved sample (*see* **Note 9**)
5. Incubate the samples on a vortex mixer at low speed at RT for 10 min. Then, centrifuge for 10 s at 18,000 × *g* to pellet the beads. Carefully transfer the supernatant to a new LoBind Eppendorf tube. Save the pelleted beads.
6. For a second round of enrichment, add fresh TiO_2 beads from the stock solution to the supernatant, this time corresponding to a bead–peptide ratio of 3:1. Repeat the procedure as described in **step 5**.
7. For a third round of enrichment, add fresh TiO_2 beads from the stock solution to the supernatant, corresponding to a bead–peptide ratio of 1.5:1. Repeat the procedure as described in **step 5** (*see* **Note 10**).
8. Pool the beads from all tubes to a new LoBind Eppendorf tube using 100 μL of Loading Buffer 1. Centrifuge as mentioned above to pellet the beads and discard the supernatant.
9. Wash the beads using 100 μL of TiO_2 Washing Buffer 1. Therefore, vortex for 15 s, centrifuge as above and carefully discard the supernatant.
10. Repeat the procedure described in **step 9** using Washing Buffer 2.
11. Dry the beads under vacuum.

12. Elute the phosphopeptides from the beads using 100 µL of TiO_2 Elution Buffer. Vortex at high speed for 15 s and incubate on a vortex mixer at low speed for 10 min at RT. Afterwards, centrifuge for 10 s at 18,000 × *g* to pellet the beads. Carefully transfer the supernatant (the eluate) to a new LoBind Eppendorf tube.
13. Add another 30 µL of Elution Buffer and vortex at high speed for 15 s. Immediately after centrifugation for 10 s at 18,000 × *g*, combine the supernatant with the eluate obtained from, **step 12**.
14. Acidify the eluate using 8 µL of 100 % FA and 2 µL of 10 % TFA (*see* **Note 11**).
15. Dry the eluate from **step 14** under vacuum.
16. Once the eluate is dried, add 1 mL of Loading Buffer 2.
17. For each 100 µg of sample, weigh in 900 µg of titanium dioxide beads. Resuspend the beads in a defined volume using Loading Buffer 2, e.g., add 45 µL of Loading Buffer 2–900 µg of TiO_2 to yield a final concentration of 20 µg/µL.
18. Repeat the same procedure as described in **steps 4–6**. However, this time only perform two rounds of TiO_2 enrichment, using bead–peptide ratios of 6:1 (first round) and 3:1 (second round), respectively.
19. Pool the beads from both tubes to a new LoBind Eppendorf tube using 100 µL of Loading Buffer 2. To pellet the beads, centrifuge for 10 s at 18,000 × *g* and discard the supernatant.
20. Wash the beads using 100 µL of TiO_2 Washing Buffer 3. Vortex for 15 s, centrifuge as above and carefully discard the supernatant.
21. Dry the beads under vacuum.
22. Elute the phosphopeptides from the beads using 100 µL of TiO_2 Elution Buffer. Vortex at high speed for 15 s and incubate the samples on a vortex mixer at low speed and RT for 10 min. Afterwards, centrifuge for 10 s at 18,000 × *g* to pellet the beads. Carefully transfer the supernatant (the eluate) to a new LoBind Eppendorf tube (*see* **Note 12**).
23. Add another 30 µL of Elution Buffer and vortex at high speed for 15 s. Immediately after centrifugation for 10 s at 18,000 × *g*, combine the supernatant with the eluate obtained from **step 22**.
24. Acidify the eluate using 8 µL of 100 % FA and 2 µL of 10 % TFA (*see* **Note 11**).

3.4 Desalting of Enriched Samples by Solid Phase Extraction

This step is important not only for desalting, but also to remove residual TiO_2 beads.

1. Prepare a C18 stage tip. Cut 5 mm from the top of a conventional 2–200 µL pipette tip. Use the rest of the tip as a blunt-ended syringe needle to stamp out a small piece of a

C18 Empore material. Use a gel loader tip to fix the C18 material on the bottom of another unaltered conventional 2–200 μL pipette tip (*see* **Note 13**).

2. Weigh in 5 mg of Oligo R3 material. Solubilize the R3 material in 200 μL of 70 % ACN.
3. Pipette 10 μL of the Oligo R3 solution into the C18 stage tip.
4. Activate the materials by pipetting 60 μL of 100 % ACN into the C18 stage tip. Pass the liquid through the materials with a filled air syringe.
5. Equilibrate the materials with 60 μL of 0.1 % TFA. Pass the liquid through the materials with a filled air syringe.
6. Load the complete eluate onto the stage tip. Save the flow-through and load once more.
7. Desalt by passing 60 μL of 0.1 % TFA through the material.
8. Elute your desalted phosphopeptides by slowly passing 60 μL of 70 % ACN, 0.1 % TFA through the stage tip. Collect the eluate in a fresh LoBind Eppendorf tube.
9. Dry the eluted sample under vacuum.

3.5 LC-MS Analysis of the Enriched Phosphopeptide Sample

1. The described setup serves as an example and can be modified according to the facilities of the laboratory as well as to the amount and complexity of the sample. However, for analyzing the subtilisin phosphopeptide sample, using high resolution and high mass accuracy for both, MS and MS/MS, is strongly recommended in order to reduce the search space for the following database search and thus increase the number of identifications at a given false discovery rate.
2. For HPLC analysis, load the samples onto the RP trap column with 0.1 % TFA (HPLC Loading Buffer) at a flow rate of 20 μL/min, followed by separation on a 50 cm RP main column using a binary gradient (HPLC Solvents A: 0.1 % FA and B: 0.1 % FA, 84 % ACN) from 3 % to 42 % Solvent B at a flow rate of 250 nL/min in 90 min, followed by 4 min at 95 % B.
3. Operate a Q Exactive Plus mass spectrometer in data-dependent acquisition mode acquiring MS survey scans at a resolution of 70,000 and a target value of 1×10^6 ions with a maximum fill time of 120 ms. Acquire MS/MS scans of the 15 most abundant ions (Top 15), using 1×10^5 ions as the target value and a maximum fill time of 250 ms. Use a normalized collision energy of 27 and a dynamic exclusion of 12 s, selecting only precursor ions with charge states between +2 and +5.

3.6 Data Analysis

1. Search the MS data against the human Uniprot database (December, 2013; 20,273 target sequences) using the Proteome Discoverer (PD) software version 1.3, as described below.

2. Process raw data using the Spectrum Selector node with default settings.
3. Use Mascot as search algorithm with the following search parameters: (1) protease "none" in case of subtilisin, or "trypsin", (2) oxidation of methionine as well as phosphorylation of serine, threonine and tyrosine as dynamic modifications, (3) carbamidomethylation of cysteine as static modification, (4) precursor mass tolerance of 10 ppm, (5) fragment ion mass tolerance of 0.02 Da.
4. Use the Peptide Validator node for false discovery rate (FDR) assessment. Apply high confidence filter corresponding to an FDR ≤1 % on the level of peptide-spectrum-matches (PSM). Additionally, filter PSM for peptide search engine rank 1.
5. Determine phosphorylation site localization probabilities using phosphoRS (version 2.0) [16]. Consider phosphoRS localization probabilities ≥ 99 % as confident.

3.7 Evaluating Digestion Reproducibility

1. For all the samples, filter only phosphopeptides that are unique for a single protein.
2. Export the PSM list from Proteome Discoverer to Microsoft Excel.
3. With the help of the ready-to-use Excel macro provided by the Mechtler group (http://ms.imp.ac.at/?goto=phosphors), determine the confident phosphorylation sites for each peptide, as well as the position of the phosphorylation within the protein sequence.
4. In Excel concatenate: (1) peptide sequence, (2) protein accession, and (3) phosphoRS phosphorylation site (only those with probabilities ≥99 %) to define unique peptides identified from the unspecific digestion with subtilisin (*see* **Notes 14** and **15**).
5. Remove all duplicates.
6. Repeat the same procedures for all replicates.
7. Plot a Venn diagram comparing the three replicates using eulerAPE [19].
8. Repeat **steps 1**–**7**, Subheading 3.7 for trypsin.
9. For each enzyme, determine how many concatenated entries (peptide sequence, protein accession, and phosphoRS phosphorylation site) overlap in at least two out of three replicates (*see* Figs. 1 and 2).

3.8 Determine Phosphorylation Sites Exclusively Obtained Using Subtilisin for Digestion

1. Proceed after **step 3**, Subheading 3.8 and concatenate (1) protein accession and (2) phosphorylation site within the protein (*see* **Note 16**).
2. Proceed as in **steps 4**–**6**, Subheading 3.7.
3. Repeat **steps 1**–**2**, Subheading 3.8 for trypsin.

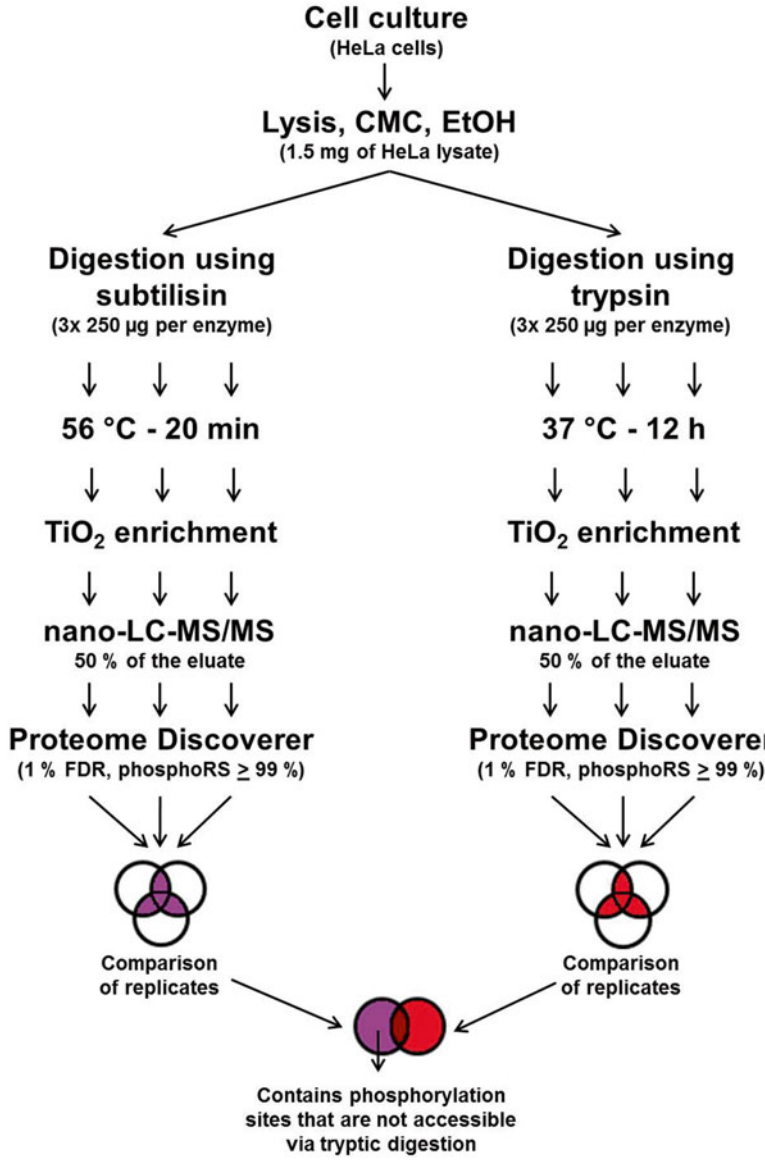


Fig. 1 Simple workflow to identify novel phosphorylation sites using subtilisin as proteolytic enzyme. To facilitate the comparison of the two digestion strategies, one HeLa stock sample was used and aliquoted before digestion. HeLa cells were lysed, carbamidomethylated, proteins precipitated with EtOH and solubilized in 2 M GuHCl buffer. Solubilized proteins were divided in a total of six aliquots (three per condition, 250 µg each) and digested either with trypsin or subtilisin. Phosphopetides were enriched with titanium dioxide (TiO_2) and half of each eluate was measured in a single nano-LC-MS/MS run. Raw files were analyzed using Proteome Discoverer 1.3 using phosphoRS 2.0 for phosphorylation site localization. Only phosphorylation sites with a probability of at least 99 % were considered. Afterwards, for both digestion strategies reproducibly identified phosphopeptides were used to compare protein phosphorylation sites that were covered with trypsin and/or subtilisin

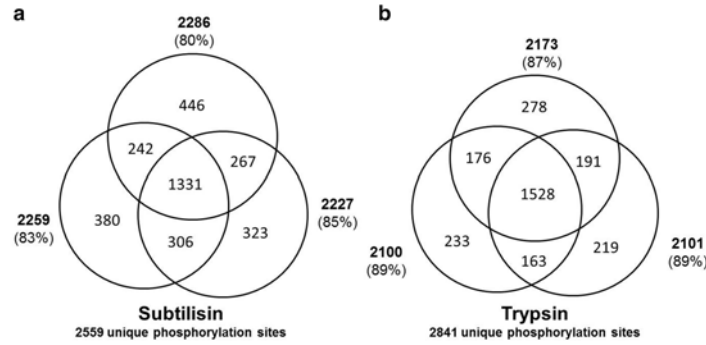


Fig. 2 Reproducibility of three technical replicates. Peptide sequence, protein accession and protein phosphorylation sites were concatenated to assess the reproducibility of **(a)** subtilisin and **(b)** trypsin digestion strategies. This is particularly important since subtilisin could yield many different peptides that contain the same phosphorylation site. For each replicate (1) the total number of phosphopeptides and (2) the percentage of phosphopeptides that have been identified in at least two out of three replicates are given, considering only phosphopeptides that pass the 1 % FDR and phosphoRS probability $\geq$ 99 % cutoffs

4. For each enzyme, determine how many concatenated entries (protein accession and phosphorylation site within the protein) overlap in at least two out of three replicates (*see* **Note 17**) (*see* Fig. 3a, b).
5. Plot two Venn diagrams as described in Subheading 3.7, **step 7**. (1) Comparing the protein phosphorylation sites identified in at least two replicates between trypsin and subtilisin (to evaluate reproducibility). (2) Comparing all identified protein phosphorylation sites between subtilisin and trypsin (to evaluate the total number of accessible phosphorylation sites when combining both enzymes) (*see* Fig. 3c, d).
6. Take the phosphorylation sites that are exclusively present in the subtilisin dataset, and retrieve a FASTA database of the corresponding protein accessions from www.uniprot.org.
7. Use the DBtoolkit [17] software for in silico digestion of the retrieved FASTA file, using trypsin as enzyme, a mass limit between 600 and 3500 Da, and a maximum of two missed cleavage sites (*see* **Note 18**).
8. In Excel, assess whether the phosphorylation sites exclusively identified in the subtilisin workflow can be covered by trypsin theoretically (i.e., are covered by the in silico generated tryptic peptides) (*see* **Note 19**).

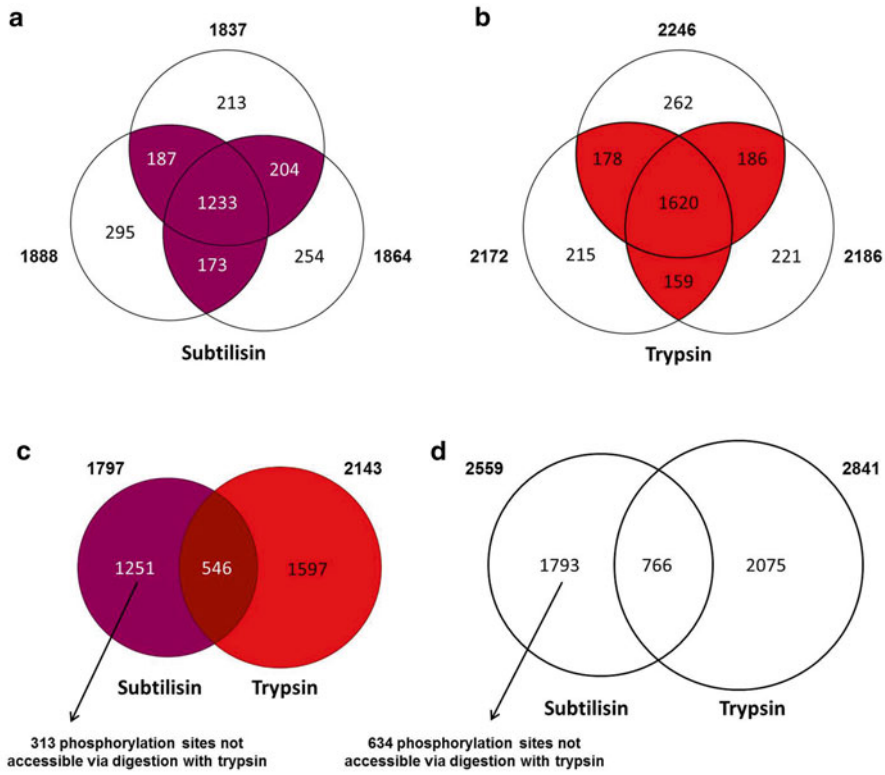


Fig. 3 Coverage of the phosphoproteome using trypsin and subtilisin. Venn diagrams depicting the overlaps of identified phosphorylation sites between the three replicates of **(a)** subtilisin and **(b)** trypsin digestion. The highlighted areas represent phosphorylation sites that have been identified in at least two out of three replicates with high confidence (1 % FDR, phosphoRS probabilities $\geq$ 99 %). Overlaps of the protein phosphorylation sites for subtilisin and trypsin, considering **(c)** only sites that were found in 2 out of 3 replicates for each digestion, and **(d)** all phosphorylation sites, respectively. Notably, the coverage of the phosphoproteome could be improved substantially. By using subtilisin the number of phosphorylation sites could be increased by 58 % **(c)** and 63 % **(d),** respectively, as compared to trypsin digestion alone. More importantly, according to in silico digestion 25 % **(c)** and 35 % **(d)** of these additional subtilisin-derived sites are theoretically inaccessible for trypsin

4 Notes

1. The cultivation of one 75 cm^2 cell culture flask of HeLa cells yields approximately 8×10^6 cells (grown to 70–80 % confluence), which corresponds to approximately 2–3 mg of protein. Any other adherent or suspension cell line is also appropriate, though few parameters such as the peptide to bead ratio might need optimization due to the potentially different phosphorylation level of cell lines.
2. The addition of phosphatase inhibitors prevents the dephosphorylation of proteins by endogenous phosphatases during sample preparation, increasing yield, reliability and reproducibility of the phosphopeptide analysis.

3. IAA is unstable and light-sensitive, therefore prepare the 1 M IAA stock freshly to minimize hydrolysis and keep it in the dark.
4. As concentrated TFA is highly corrosive, wear protective clothing (lab coat, goggles and gloves) and work under a chemical fume hood.
5. As the loading step is essential for the success of the phosphopeptide enrichment, we recommend preparing the TiO_2 Loading Buffer freshly each time.
6. For the determination of the protein concentration perform a colorimetric assay, such as BCA protein assay, Bradford or Lowry. Accurate determination is important for subsequent phosphopeptide enrichment and total amount of proteolytic enzyme and TiO_2 beads.
7. Though sequencing grade modified trypsin retains part of its activity under denaturing conditions up to 2 M GuHCl (see Promega usage information), dilute the Lysis Buffer to a final concentration of 0.2 M GuHCl prior to protein digestion to ensure full enzymatic activity.
8. In contrast to C18 RP chromatography, the monolithic system enables the detection of peptides and proteins in a single LC run and is more robust against detergents. Compared to digestion controls using SDS-PAGE and silver staining, the monolithic separation is 5–10× more sensitive and faster. In addition, the UV traces allow for a better comparison of digestion reproducibility [5].
9. Note that the TiO_2 suspension needs to be mixed well to ensure equal bead–peptide ratios for all samples.
10. For complex samples a third incubation step is recommended. In our hands usually up to 5 % more phosphopeptides can be identified by doing so.
11. To prevent loss of the phosphate moiety due to β-elimination under highly alkaline conditions, directly acidify the phosphopeptide sample after the elution step.
12. Despite the removal of residual beads during the following desalting step, try to transfer as few beads as possible using a GELoader tip.
13. A more detailed description to build a self-made StageTip is described in Rappsilber et al. [20].
14. In Excel, applying the Excel function “concatenate”, place a semicolon in between the values to concatenate: (1) peptide sequence, (2) protein accession, and (3) the confident phosphorylation sites obtained from the ready-to-use macro Excel sheet provided by the Mechtler group, e.g., ‘SAPASPTHPGLMSPR; P85037; S416; S420; S428’. Afterwards, remove duplicates. Compare the concatenations

Table 1
Assessment of phosphorylation sites which are not accessible via digestion with trypsin

Accession	In silico generated tryptic peptides (first—last AA)	Phosphorylation sites only identified after subtilisin digestion	Accessible via tryptic digestion?
P05787	33–47	43	Yes
P05787	41–47	74	No
P05787	89–96		
P05787	89–101		

According to DBtoolkit, a tryptic digestion of Keratin, type II cytoskeletal 8 (P05787) generates a total of 145 different tryptic peptides with a molecular weight between 600 and 3500 Da and a maximum of two missed cleavages. Notably, the phosphorylation site Ser74, identified only after subtilisin digestion, is not covered by any of the in silico generated tryptic peptides

from each replicate using a Venn diagram. To compare the phosphorylation sites obtained with subtilisin and trypsin, concatenate only the (1) protein accession and (2) phosphorylation site within the protein.

15. Only phosphorylation sites with a probability equal or higher than 99 % are considered. In any case, the phosphorylation probability cutoff can be altered according to one's own requirements (Loroch et al. 2014, submitted).
16. Note that subtilisin and trypsin digestions may yield different peptide sequences for the same phosphorylation site. To compare phosphorylation sites between subtilisin and trypsin, it is therefore necessary to concatenate (1) protein accession and (2) phosphorylation site within the protein.
17. To assess reproducibility of the digestion workflow, consider only concatenations that appear in at least two out of three replicates.
18. The selected mass range threshold covers what is expected in a typical trypsin-based shotgun experiment.
19. Use the DBtoolkit-generated in silico digested peptide FASTA database to assess whether a phosphorylation site identified after subtilisin digestion can theoretically be covered by trypsin digestion (*see* Table 1).

Acknowledgements

The financial support by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and by the CAPES Foundation is gratefully acknowledged.

References

1. Loroch S et al (2013) Phosphoproteomics—More than meets the eye. Electrophoresis 34(11):1483–1492
2. Mann M et al (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20(6):261–268
3. Yan JX et al (1998) Protein phosphorylation: technologies for the identification of phosphoamino acids. J Chromatogr A 808(1–2):23–41
4. Sickmann A, Meyer HE (2001) Phosphoamino acid analysis. Proteomics 1(2):200–206
5. Burkhart JM et al (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. J Proteomics 75(4): 1454–1462
6. Picotti P, Aebersold R, Domon B (2007) The Implications of Proteolytic Background for Shotgun Proteomics. Mol Cell Proteomics 6(9):1589–1598
7. Proc JL et al (2010) A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. J Proteome Res 9(10):5422–5437
8. Olsen JV, Ong S-E, Mann M (2004) Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. Mol Cell Proteomics 3(6):608–614
9. Bunkenborg J, Espadas G, Molina H (2013) Cutting Edge Proteomics: Benchmarking of Six Commercial Trypsins. J Proteome Res 12(8):3631–3641
10. Vandermarliere E, Mueller M, Martens L (2013) Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom Rev 32(6):453–465
11. Walmsley SJ et al (2013) Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin As a Significant Source of Variability in Proteomics. J Proteome Res 12(12):5666–5680
12. Dickhut C et al (2014) Impact of Digestion Conditions on Phosphoproteomics. J Proteome Res 13(6):2761–2770
13. Hamady M et al (2005) Does protein structure influence trypsin miscleavage? Engineering in Medicine and Biology Magazine. IEEE 24(3): 58–66
14. Gilmore J, Kettenbach A, Gerber S (2012) Increasing phosphoproteomic coverage through sequential digestion by complementary proteases. Anal Bioanal Chem 402(2): 711–720
15. Jacobs M et al (1985) Cloning, sequencing and expression of subtilisin Carlsberg from Bacillus licheniformis. Nucleic Acids Res 13(24): 8913–8926
16. Taus T et al (2011) Universal and Confident Phosphorylation Site Localization Using phosphoRS. J Proteome Res 10(12): 5354–5362
17. Martens L, Vandekerckhove J, Gevaert K (2005) DBToolkit: processing protein databases for peptide-centric proteomics. Bioinformatics 21(17):3584–3585
18. Palmisano G et al (2012) A Novel Method for the Simultaneous Enrichment, Identification, and Quantification of Phosphopeptides and Sialylated Glycopeptides Applied to a Temporal Profile of Mouse Brain Development. Mol Cell Proteomics 11(11):1191–1202
19. Micallef L, Rodgers P (2014) eulerape: drawing area-proportional 3-Venn diagrams using ellipses. PLoS One 9(7):e101717

Chapter 16

Online LC-FAIMS-MS/MS for the Analysis of Phosphorylation in Proteins

Hongyan Zhao, Andrew J. Creese, and Helen J. Cooper

Abstract

High-field asymmetric waveform ion mobility spectrometry (FAIMS) is a gas-phase separation technique which, when coupled with liquid chromatography tandem mass spectrometry, offers benefits for analysis of complex proteomics samples such as those encountered in phosphoproteomics experiments. Results from LC-FAIMS-MS/MS are typically complementary, in terms of proteome coverage and isomer identification, to those obtained by use of solution-phase separation methods, such as prefractionation with strong cation-exchange chromatography. Here, we describe the protocol for large-scale phosphorylation analysis by LC-FAIMS-MS/MS.

Key words Phosphorylation, Ion mobility spectrometry, FAIMS

1 Introduction

High-field asymmetric waveform ion mobility spectrometry (FAIMS), also known as differential ion mobility, is a technique in which gas-phase ions are separated based on their ion mobilities in an asymmetric electric field. The ions are infused between two electrodes and are exposed to alternating high and low electric fields in which they have different ion mobilities. Only the ions with "balanced" motilities between the high and low electric fields will exit the device, while all others will drift towards one of the two electrodes. By applying a compensation voltage (CV) to one of the electrodes, it is possible to correct for this drift. Scanning the compensation voltage allows selective transmission of ions according to their differential mobilities [1, 2]. The principles of FAIMS separation are illustrated in Fig. 1.

FAIMS offers several benefits for mass spectrometry-based proteomics research. FAIMS can maximize the proportion of multiply-charged ions while minimizing interfering singly charged species, which is advantageous for both collision-induced dissociation (CID) and electron transfer dissociation (ETD) mass

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_16,

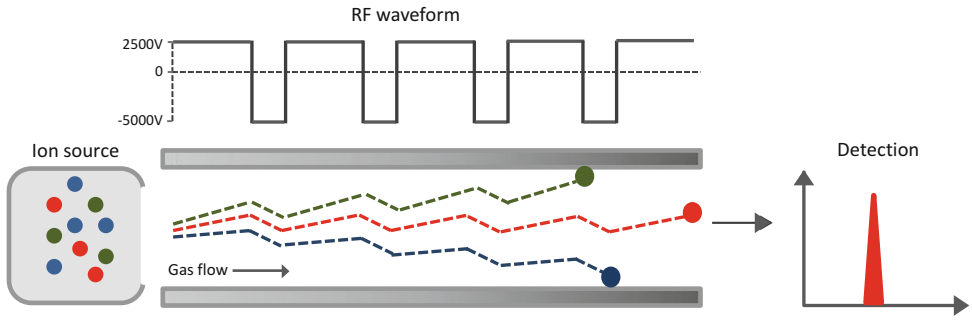


Fig. 1 Schematic diagram of FAIMS separation

spectrometry. A number of groups have demonstrated that FAIMS enhances sensitivity by reducing background contaminants thus improving signal-to-noise ratios and facilitating peak detection [3, 4]. Application of FAIMS in phosphoproteomics workflows has resulted in increased coverage of the phosphoproteome [5]. Creese et al. [6] demonstrated that liquid chromatography (LC)-FAIMS-MS/MS outperformed LC-MS/MS in the identification of isomeric phosphopeptides from a phosphopeptide library.

The challenge for large-scale phosphorylation studies is sample complexity. Samples need to go through the following steps before mass spectrometry analysis: (a) trypsin digestion; (b) fractionation, for example, with strong cation exchange (SCX) or gel electrophoresis; and (c) phosphoenrichment [7]. A complementary approach to liquid phase separation (such as SCX) for overcoming sample complexity is gas-phase fractionation by FAIMS [8] (a consequence of gas-phase fractionation is that the workflow order is changed: FAIMS fractionation takes place after phosphoenrichment). In this approach, the cell lysate is digested with trypsin and the resulting peptide mixture is enriched for phosphopeptides. The sample is subjected to multiple LC-FAIMS-MS/MS analyses at different, and constant, compensation voltages (ranging from −20 V to −50 V).

2 Materials

2.1 Materials for Sample Preparation

1. 1 mg/ml stock solution of Substance P in water. Further diluted to 2 pmol/μL with 30 % acetonitrile, 0.1 % formic acid.
2. 100 mM ammonium bicarbonate.
3. Dionex™ Protein Mixture Digest dissolved in 0.1 % formic acid to give a concentration of 50 fmol/μL.
4. 8 mg of HEK 293T (or other) cell lysate in cell lysis buffer (1M Tris–HCl, 10 % Triton-100, 5M NaCl).

5. Trypsin Gold (mass spectrometry grade) dissolved in 50 mM acetic acid to a concentration of 1 μg/μL.
6. Sep-Pak Plus Light Cartridge, Titansphere™ Phos-TiO_2 kit and Ziptip Pipette tips C18.
7. 0.1 % formic acid.
8. Trifluoroacetic acid (TFA).
9. Acetic acid.
10. 5 % ammonium hydroxide.
11. 5 % pyrrolidine.

2.2 Mass Spectrometer and Software

1. LTQ Orbitrap Velos™ ETD mass spectrometer (Thermo Fisher Scientific).
2. FAIMS device (Thermo Fisher Scientific).
3. HESI–II probe (Thermo Fisher Scientific), modified as described by Swearingen et al. [9], and a PicoTip™ emitter, OD 360 μm and ID 20 μm (New Objective).
4. Air Compressor 4000-40M (Jun-air). The total gas flow needs to be maintained at approximately 86 L/min.
5. Data acquisition software: Xcalibur 2.1 (Thermo Fisher Scientific).
6. Search software: Proteome Discoverer 1.3 (Thermo Fisher Scientific).

2.3 Reversed-Phase Liquid Chromatography (RPLC)

1. UltiMate NCS-3500RS binary pump (Dionex) equipped with an UltiMate WPS3000 autosampler (Dionex).
2. Acclaim Pepmap100 15 cm × 75 μm C18 analytical column and a 2 cm × 75 μm trap column, both packed with 3 μm C18 particles.
3. LC buffers: Mobile phase A 0.1 % formic acid in H_2O and mobile phase B 0.1 % formic acid in acetonitrile.
4. Software: Dionex Chromatography MS links V6.8.

3 Methods

3.1 Sample Preparation

1. Digest the cell lysate with 1 μg/μL trypsin in 100 mM ammonium bicarbonate at 37 °C overnight at a protease:protein ratio of 1:100.
2. Add 10 μL TFA to 2 mL of protein digest sample.
3. Desalt the sample by use of Sep-Pak C18 cartridges: (a) wash cartridges with 4 mL of acetonitrile; (b) condition cartridge with 1.5 mL of 50 % acetonitrile/0.5 % acetic acid; (c) equilibrate with 4 mL of 0.1 % TFA; (d) load sample onto cartridge; (e) wash the cartridge with 4 mL of 0.1 % TFA followed by

1.5 mL of 0.5 % acetic acid; (f) elute peptides from cartridge with 2 mL of 50 % acetonitrile/0.5 % acetic acid. Dry the sample by vacuum centrifugation.

4. Split the sample into 13 equal fractions. Enrich each fraction for phosphorylated peptides by use of a Phos-TiO_2 kit (*see* **Note 1**). Prepare 3.0 mL of Buffer A (600 μL of 2 % TFA: 2.4 mL of acetonitrile) and 2.0 mL of Buffer B (0.5 mL lactic acid: 1.5 mL Buffer A). Connect a "Centrifugal Adaptor" to a "Waste Fluid Tube" and insert a "Spin Tip" into the "Centrifugal Adaptor." Add 20 μL of Buffer A to Spin Tip, and centrifuge (3000 × *g*, 2 min). Add 20 μL of Buffer B to "Spin Tip," and centrifuge (3000 × *g*, 2 min). Remove flow-through from "Waste Fluid Tube." Resuspend sample in 50 μL Buffer B. Load sample into "Spin Tip," and centrifuge (1000 × *g*, 10 min). Reload sample from "Waste Flow Tube" into "Spin Tip" and centrifuge (1000 × *g*, 10 min). Add 20 μL of Buffer B and centrifuge (3000 × *g*, 2 min). Add 20 μL of Buffer A and centrifuge (3000 × *g*, 2 min), and repeat three times. Remove the "Spin Tip" and "Centrifugal Adaptor" and place into a "Recovery Tube." Add 50 μL of 5 % ammonium hydroxide solution to "Spin Tip" and centrifuge (1000 × *g*, 5 min). Add 50 μL of 5 % pyrrolidine solution to the Spin Tip and centrifuge (1000 × *g*, 5 min). Remove the supernatant by vacuum centrifugation.
5. Desalt each fraction with Ziptip Pipette tip. Resuspend each fraction in 10 μL of 0.1 % TFA. First, wet Ziptip with 10 μL of acetonitrile (repeat once). Aspirate 10 μL of 0.1 % TFA solution and dispense to waste (repeat twice). Bind peptides by aspirating and dispensing the sample seven to ten times for maximum binding. Wash with 10 μL of 0.1 % TFA and dispense to waste (repeat twice). Elute peptides with 10 μL of 70 % acetonitrile in 0.1 % TFA and dispense into a clean vial. Remove the supernatant by vacuum centrifugation.
6. Resuspend the phosphopeptides in 10 μL 0.1 % formic acid.

3.2 Tune File Creation

1. Open LTQ Tune.
2. Open a Tune file for standard LC-MS/MS analysis.
3. Click on the "Define Scan" icon to set the parameters as follows:

Scan mode	Settings	Scan mode	Settings
Analyzer	FTMS	Micro scan	1
Mass range	High	Max. Inject Time (ms)	1000
Resolution	60000	Tube lens voltage (V)	100
Scan type	Full	Scan Ranges	380–2000

4. HESI–II settings: Click on the "Source" icon to set the parameters as follows:

HESI–II parameters	Values	HESI–II parameters	Values
Sheath gas flow rate (arb)	1~3	Capillary temperature (°C)	250
Aux/sweep gas flow rate (arb)	0	Capillary voltage (V)	25
Spray voltage (kV)	2.5~3.0	Tube lens voltage (V)	100

5. FAIMS settings: Click on the "FAIMS" icon to set the parameters as follows:

FAIMS parameters (V)	Values	FAIMS parameters	Values
Compensation voltage	−25	Total gas flow (L/min)	2.9
Outer bias voltage	0	He %	50
Dispersion voltage	−5000	Inner/outer electrode temperature (°C)	70/90

6. ETD settings:
   (a) Choose "Reagent Ion Source" to open the Reagent Ion Source dialog box. Check the boxes "Reagent Ion Source On," "Filament On," and "View Reagent Ion Spectra." Click OK.
   (b) Click on the "Tune" icon. Select the Automatic tab and click start. The system starts automatically tuning the reagent ion optics. After completing tuning, the message "Optimization Complete" will be displayed. If ETD reagent ion signal is below 5.0 E6, repeat the optimization.
   (c) Click "Save As" to save the file with a new name.

3.3 LC-FAIMS-MS/MS Method

To facilitate the identification of PTMs, we used a "Top-7" mass spectrometry method, i.e., a full FT-MS survey scan followed by seven MS/MS scans in which the seven most abundant precursor ions were fragmented. The FAIMS method is the "external stepping" method [8] in which multiple LC MS/MS analyses are performed each at a separate, and constant, compensation voltage (*see* **Note 2**).

1. In Xcalibur, choose "Instrument Setup" and "Data Dependent MS/MS." An instrument setup window will be displayed.
2. LC settings: click on the "Dionex" icon and use Wizard to set up the LC settings.

   Under "Valve Time," change the valve from 6_1 (trap and column out of line) to 1_2 (trap and column in line) at 6 min and switch it back at 50 min. Under "Gradient Type," set up a

30-min gradient from 3.2 % to 44 % mobile phase B, followed by a 10-min wash with 90 % mobile phase B and re-equilibration (15 min) with 3.2 % mobile phase B. Use default settings for the rest of the parameters.

3. LTQ settings: Click on the "LTQ" icon and change the Tune method to the Tune file created in the previous section.
4. Change the number of scan events to 8. Scan events 2–8 are dependent scans (*see* **Note 3**).
5. For dependent scan settings, go to "Scan event and Activation." Set the parameters as follows:

Scan description	Settings	Dependent scan	Settings
Analyzer	FTMS	Activation type	ETD
Resolution	30,000	Default charge state	2
Scan type	Full	Isolation width (*m/z*)	2.0
Data type	Profile	Activation time (ms)	100.00

6. Change the CV (bottom left corner) accordingly (–20 V to –50 V, in 2.5 V steps). Save each method individually.

3.4 Configuration of FAIMS

1. Connect the FAIMS waveform generator and temperature control module to the mass spectrometer. Connect all the cables and gas lines to FAIMS interface, i.e., high-voltage lead, bias-voltage cable, dispersion voltage cable, carrier gas line, and heating gas lines.
2. Open the helium canister regulator. Adjust to a pressure of 60 psi.
3. Turn on the air compressor and adjust the pressure to 50 psi.
4. Close all computer programs and open Xcalibur Instrument Configuration.
5. Select the "LTQ," and click on the "configure" box to display the "LTQ Configuration" box.
6. Select "FAIMS" and tick the "FAIMS Configured" box.
7. Click "OK" and restart computer to apply the settings.
8. When Instrument Console appears, press the "Reset" button on LTQ to restart mass spectrometer. The instrument is now in standby mode and ready to use.

3.5 Direct Infusion

Before sample analysis, the FAIMS system is optimized by infusing a standard peptide, in our case Substance P (2 pmol/μL). Firstly, a CV scan is performed to determine the optimum CV for Substance P. Further optimization can then be carried out to determine the FAIMS settings and the optimum position of the probe (*see* **Note 4**).

3.5.1 Syringe Pump Setting

1. Fill the syringe with Substance P solution. Assemble the syringe pump by connecting the LC union (finger-tight fitting) and the syringe.
2. Place the syringe into the syringe holder.

3.5.2 HESI–II Probe

1. Insert the HESI–II probe to a depth of "D" position (indicated by the markers on the probe).
2. Connect the sheath gas line, auxiliary gas line and 8 kV voltage cable.
3. The vaporizer cable is connected with the interlock socket; therefore the heating capability is not utilized in the FAIMS experiment.
4. Connect HESI–II probe to the syringe pump.

3.5.3 CV Optimisation

1. Click on the "Start" icon to start the FT-MS scan.
2. Set the CV to −30 V for optimum intensity of signal of Substance P (*see* **Note 5**). Substance P is introduced at a flow rate 0.35 μL/min. Allow 3–5 min for the sample to flow to the ion source.
3. Click on the "Tune" icon and select the FAIMS tab. In "Masses," type in the *m/z* of the species under investigation. In this case, use the 2^+ ions of Substance P: *m/z* 674.36. Set the CV scan range from −50 V to −10 V. Click "Start" to start CV scanning.
4. Click on the "Display Graph" button: the CV scans of ion *m/z* 674.36 will be displayed.
5. When the CV scan has finished, click "Accept" to apply the optimum CV for further signal optimization.

3.5.4 Optimization of FAIMS and Other Parameters

To maximize the ion signal, spray voltage and gas composition can be adjusted accordingly. The position of capillary needle also plays an important role in determining peak capacity. The optimum signal is observed when the tip of the needle is approximately 5 mm away and upper right from the orifice (*see* **Notes 6** and **7**).

3.6 LC-FAIMS-MS/MS Analysis

1. Connect HESI–II probe with the LC column.
2. To evaluate the performance of LC-FAIMS-MS/MS, Dionex™ Protein Mixture Digest is used. Place 3–4 μL of the sample in a well in a 96-well microtiter plate and place the plate in the autosampler.
3. Go to Xcalibur and click on the "Sequence Setup" icon. To create a new row, fill in the name, path, position on the microtiter plate, and volume (2 μL) of the sample. Double-click on the "Inst Meth" box to select the method created in Subheading 3.3 and save the sequence.

4. Click on the sequence created in previous step, and then click on the "Start" icon to start the sequence.
5. Search the resulting dataset against database containing the six proteins (manually created containing 691 proteins) using SEQUEST in Proteome Discoverer 1.3. The database is available for download from http://www.biosciences-labs.bham.ac.uk/cooper/downloads/sixmix.zip. The expected protein coverage for each protein is between 50 and 70 %. Under optimal conditions, the intensity of highest peak observed with FAIMS should be about 20 % of that obtained with LC-MS/MS while the protein coverage is similar.
6. To perform the LC-FAIMS-MS/MS analyses for the phosphopeptide samples, load the thirteen 10 μL samples into wells in a 96-well microtiter plate and place the plate in the autosampler.
7. Go to Xcalibur and click on the "Sequence Setup" icon. Create new rows for each sample with different names. Fill in the name, path, position on the microtiter plate, and volume (9 μL) of the sample. Double click on the "Inst Meth" box to select the method created in Subheading 3.3, **step 6** (according to applied CV) and save the sequence.
8. Select the sequence (all rows) created in previous step, and then click on the "Start" icon to start the sequence.
9. Search the resulting dataset against an appropriate protein database using your preferred search algorithm (*see* **Note 8**).

4 Notes

1. The phosphoenrichment step prior to FAIMS analysis is necessary for phosphopeptide identification. As no fractionation is performed before phosphoenrichment, care should be taken to ensure the enrichment procedure is efficient. It may be necessary to optimize the amount of protein against the number of tips for different samples.
2. The external CV stepping method, in which the CV remains constant throughout each LC-FAIMS-MS/MS analysis, is used in this assay. Alternatively, an internal CV stepping method can be used. In the internal CV stepping method, the mass spectrometer performs a survey scan at CV of, e.g., –25 V and subsequent ETD event(s) at the same CV value. The sequence then cycles through CV of –25 V, –30 V…–50 V. In our hands, however, the internal stepping method yields fewer identifications [8].

3. The number of scan events in this method is eight: one survey scan and seven subsequent ETD events. Different numbers of scan events can be used according to mass spectrometer performance and sample complexity. The alternating CID-ETD fragmentation method is also suitable for PTM identification.
4. Based on observations in our laboratory, the sensitivity of FAIMS is about 1/8th–1/5th of that observed with non-FAIMS analysis. However, with FAIMS the increased signal-to-noise (S/N) ratio should compensate for the signal loss, which in turn will improve the overall performance of FAIMS.
5. When optimizing the performance of FAIMS, a starting CV value needs to be used before tuning for optimum CV. In this case, we used −30 V. It is possible to use any CV from −10 V to −50 V provided that the peak at the target *m/z* can be detected.
6. When optimizing the performance of FAIMS, if the initial signal is weak and unstable, the following parameters can be adjusted: spray voltage, carrier gas composition and sheath gas flow rate. Sheath gas is used to assist ions focusing into the entrance orifice. For example, when the spray needle is far from the orifice a higher flow rate would be required. Improved results were observed when the flow rate of sheath gas was 2 or 3 (arbitrary units).
7. If the FAIMS ion signal is lower than expected, check whether the FAIMS electrodes are clean. Normally, the FAIMS electrodes need to be cleaned every 1 or 2 weeks if the FAIMS device is used frequently.
8. If using HEK 293T cells, search the data against SWISS-PROT human database using SEQUEST and Mascot algorithms in Proteome Discoverer 1.3.

Acknowledgments

The Advion Triversa Nanomate, Dionex LC and Thermo Fisher Velos Orbitrap mass spectrometer used in this research were funded through the Birmingham Science City Translational Medicine: Experimental Medicine Network of Excellence Project, with support from Advantage West Midlands (AWM). The Chinese Scholarship Council is gratefully acknowledged for funding. HJC is an EPSRC Established Career Fellow (EP/L023490/1).

References

1. Swearingen KE, Moritz RL (2012) High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. Expert Rev Proteomics 9:505–517
2. Guevremont R (2004) High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. J Chromatogr A 1058:3–19
3. Venne K, Bonneil E, Eng K et al (2005) Improvement in peptide detection for proteomics analyses using nanoLC-MS and high-field asymmetry waveform ion mobility mass spectrometry. Anal Chem 77:2176–2186
4. Saba J, Bonneil E, Pomies C et al (2009) Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear Ion trap/orbitrap mass spectrometer. J Proteome Res 8:3355–3366
5. Bridon GL, Bonneil E, Muratore-Schroeder T et al (2011) Improvement of phosphoproteome analyses using FAIMS and decision tree fragmentation. Application to the insulin signaling pathway in Drosophila melanogaster S2 cells. J Proteome Res 11:927–940
6. Creese AJ, Smart J, Cooper HJ (2013) Large-scale analysis of peptide sequence variants: the case for high field asymmetric waveform ion mobility spectrometry. Anal Chem 85:4836–4843
7. Villén J, Gygi SP (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3:1630–1638
8. Creese AJ, Shimwell NJ, Larkins KP et al (2013) Probing the complementarity of FAIMS and strong cation exchange chromatography in shotgun proteomics. J Am Soc Mass Spectrom 24:431–443
9. Swearingen KE, Hoopmann MR, Johnson RS et al. (2012) Nanospray FAIMS fractionation provides significant increases in proteome coverage of unfractionated complex protein digests. Mol Cell Proteomics 11 doi:10.1074/m111.014985

Chapter 17

Simple and Reproducible Sample Preparation for Single-Shot Phosphoproteomics with High Sensitivity

Rosa R. Jersie-Christensen, Abida Sultan, and Jesper V. Olsen

Abstract

The traditional sample preparation workflow for mass spectrometry (MS)-based phosphoproteomics is time consuming and usually requires multiple steps, e.g., lysis, protein precipitation, reduction, alkylation, digestion, fractionation, and phosphopeptide enrichment. Each step can introduce chemical artifacts, in vitro protein and peptide modifications, and contaminations. Those often result in sample loss and affect the sensitivity, dynamic range and accuracy of the mass spectrometric analysis. Here we describe a simple and reproducible phosphoproteomics protocol, where lysis, denaturation, reduction, and alkylation are performed in a single step, thus reducing sample loss and increasing reproducibility. Moreover, unlike standard cell lysis procedures the cell harvesting is performed at high temperatures (99 °C) and without detergents and subsequent need for protein precipitation. Phosphopeptides are enriched using TiO_2 beads and the orbitrap mass spectrometer is operated in a sensitive mode with higher energy collisional dissociation (HCD).

Key words Phosphoproteomics, Phosphopeptide enrichment, Sample preparation, Heated guanidinium chloride lysis, Mass spectrometry

1 Introduction

Mass spectrometry (MS)-based proteomics has emerged as a robust, sensitive, and reliable technology for global characterization of posttranslational modifications (PTMs) of proteins and peptides. Reversible phosphorylation is a ubiquitous PTM and important in defining functional characteristics of a number of proteins involved in virtually all cellular signaling networks, such as signal transduction, cell proliferation, differentiation, apoptosis, and metabolism [1, 2]. Site-specific phosphorylation of proteins is a transient and dynamic event and phosphorylation sites on proteins are often of low stoichiometry and therefore of relatively low abundance [3]. As a result, efficient sample preparation and phosphopeptide enrichment strategies are essential for successful phosphoproteomics studies. In-depth phosphoproteome analysis by

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_17,

quantitative mass spectrometry often requires large amounts of starting protein material, extensive fractionation and many hours or days of MS acquisition time, increasing costs and limiting the number of samples that can be measured. The present chapter describes a simplified and robust sample preparation procedure for phosphopeptide enrichment without any fractionation, allowing for comprehensive, in-depth phosphoproteome coverage of up to 10,000 unique phosphopeptides from 1 mg of protein starting material in a single 4 h LC-MS run on a Q-Exactive mass spectrometer. The following protocol has been used with reproducible success down to 10 µg of starting material generating up to 1000 unique phosphopeptides (*see* Fig. 1). For details about ultra-deep phosphoproteome coverage enabled by off-line peptide fractionation the reader is referred to Chapters 10, 11, and 12.

In traditional cell lysis protocols every step of the procedure is performed cold to minimize sample preparation-induced or stress-related alterations of the proteome and protein modification profile. At low temperatures most of the cellular activity is halted preserving the in vivo state of the cells. An alternative way of quenching all cellular activity is by immediate heat denaturation of

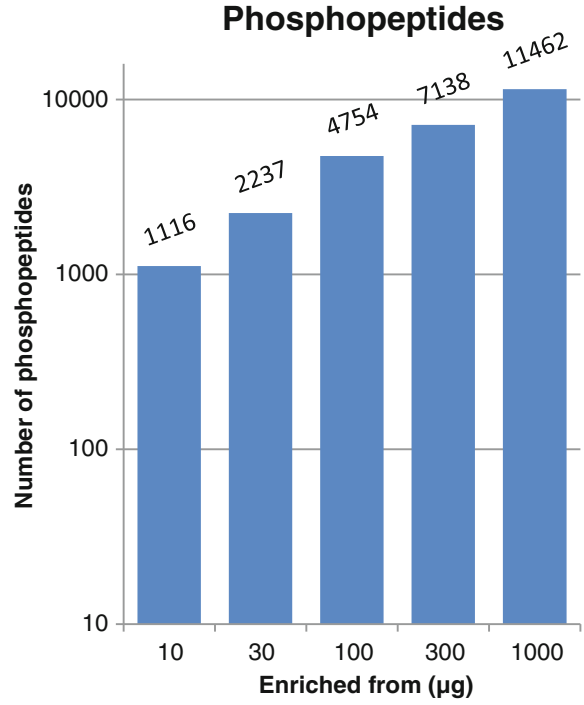


Fig. 1 A dilution series over varying amounts of protein starting material (10 µg–1 mg) was subjected to phosphopeptide enrichments to establish a measure of sensitivity of the protocol

all proteins, which can be achieved by adding hot lysis buffer to the cells. Lysis buffers are typically detergent based. Dialysis or precipitation of the proteins makes it possible to remove the detergents, which are incompatible with peptide separation on reversed-phase C_{18} columns and strongly interfere with peptide ionization in the electrospray source of the mass spectrometer. To minimize sample losses by protein precipitation lysis buffers without detergents, but instead made up of chaotropic agents such as urea and guanidinium hydrochloride (GuHCl), can be used to disrupt the cell membrane [4]. Urea can decompose and form isocyanic acid, which reacts with the primary amines of the protein and peptides. Such carbamylation of peptides will cause increased sample complexity and reduced peptide identification. This process is accelerated at elevated temperatures or slightly acidic conditions [5]; thus the chaotrope of choice for heated lysis buffer is GuHCl [6]. More importantly, when comparing traditional RIPA lysis [7] with boiling GuHCl, we noticed a significant decrease in the number of in vitro artifacts, introduced during sample preparation, with 15 % of the identified peptides modified by in vitro artifacts after RIPA lysis and only 5 % after lysis in warm GuHCl buffer. As expected we also see a higher recovery of protein amounts using the GuHCl lysis procedure.

In the traditional proteomics sample preparation workflow of protein digestion, lysis is often followed by a cysteine disulfide bridge reduction step with dithiothreitol (DTT) and a subsequent alkylation step with either chloroacetamide (CAA) or iodoacetamide (IAA) to prevent reformation of disulfide bonds. These two steps cannot be combined as IAA will react with the thiol groups on DTT and thereby inhibit the reducing effect of the DTT. Using tris(2-carboxyethyl)phosphine (TCEP) as a reducing agent instead of DTT allows for simultaneous reduction and alkylation as TCEP does not contain any thiol groups and will not react with CAA or IAA. In addition, when using GuHCl for lysis, this mix of reducing and alkylating agent can be added directly to the lysis buffer enabling lysis, reduction, and alkylation in one step and thus significantly simplifying the protocol and reducing the time spent on sample preparation (*see* Fig. 2).

Analyzing peptides in complex mixtures by mass spectrometry (MS) is always a compromise between quality and quantity of tandem mass spectra (MS/MS) and MS running time, and for typical proteome samples analyzed on an orbitrap-type instrument the most critical feature and limiting factor is the scan speed of the instrument [8]. However, in phosphoproteomics there are additional challenges in the MS/MS acquisition; apart from identification of phosphopeptides it is also necessary to accurately localize the phosphorylation site(s) with single-amino acid resolution. To achieve this, it is a necessity to generate high-quality MS/MS spectra, which requires longer acquisition scan times. Thus, for

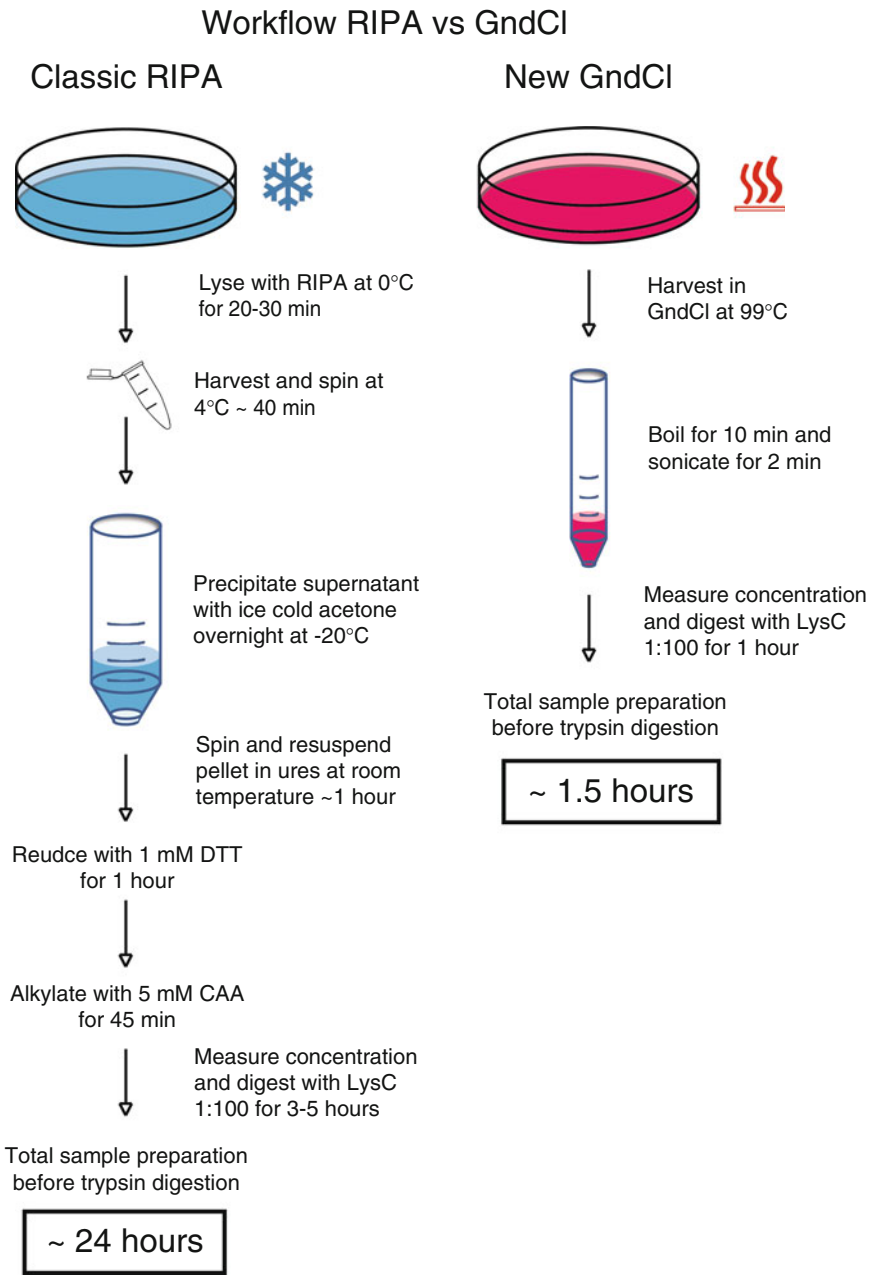


Fig. 2 Comparison of two sample lysis methods for phosphopeptide enrichments. The RIPA lysis protocol comprises multiple steps including lysis, precipitation, reduction, and alkylation, while guanidinium hydrochloride lysis is performed in a single step, where the cells are harvested and lysed in heated GuHCl buffer containing both the reducing and alkylating agents. The extracted proteins are subjected to LysC and tryptic digestions followed by phosphopeptide enrichment

comprehensive analysis of phosphoproteomes longer scan cycles are required than for typical proteome samples, for which faster scanning methods are usually sufficient for peptide identification.

2 Materials

All buffers are made with sequencing grade chemicals and ultrapure water (Milli-Q).

2.1 Cell Lysis

Experiments were performed using the adherent HeLa cervix carcinoma cell line grown in Dulbecco's modified Eagle's medium (DMEM) with L-glutamine, supplemented with 10 % fetal bovine serum (FBS) and 100 U/ml penicillin/streptomycin. Use appropriate media for your cell line of choice.

1. Human cervical epithelial cancer cell line HeLa or similar.
2. Sterile petri dishes (150 mm).
3. Lysis buffer: 6 M Guanidinium hydrochloride (GuHCl), 5 mM tris(2-carboxyethyl)phosphine (TCEP), 10 mM chloroacetamide (CAA), 100 mM Tris–HCl pH 8.5 (*see* **Note 1**).
4. Phosphate-buffered saline (PBS).
5. Rubber policeman cell scraper.
6. Heat block for centrifuge/microcentrifuge tubes with heating up to 99 °C.
7. Probe sonicator, Sonics Vibra Cell (VCX130).
8. Quick Start™ Bradford 1× Dye Reagent (Bio-Rad) or other Bradford- or BCA-based protein concentration quantification assay.

2.2 Protein Digestion

1. Lysyl endopeptidase (LysC), stock solution 0.5 mg/ml in 50 mM ammonium bicarbonate.
2. Digestion buffer: 25 mM Tris–HCl pH 8.5.
3. Modified trypsin (Sigma-Aldrich), stock solution 0.5 mg/ml in 50 mM acetic acid.
4. Trifluoroacetic acid (TFA).

2.3 Peptide Desalting

1. Reversed-phase C_{18} Sep-Pak cartridge (*see* **Note 2**).
2. Wash buffer I: 100 % acetonitrile (ACN).
3. Wash buffer II: 0.1 % TFA.
4. Elution buffer I: 40 % ACN in 0.1 % TFA.
5. Elution buffer II: 60 % ACN in 0.1 % TFA.
6. NanoDrop 2000 spectrophotometer (Thermo Scientific).
7. SpeedVac concentrator.

2.4 Phosphopeptide Enrichment

1. Reversed-phase C_8 (3 M Empore disk) single-layered StageTips (one per sample per incubation) (*see* **Note 3**).
2. Reversed-phase C_{18} (3 M Empore disk) double-layered StageTips (one per sample per incubation) (*see* **Note 3**).
3. 5 µm TiO_2 beads (Titansphere, GL Sciences) solution in 20 mg/ml 2,5-dihydroxybenzoic acid (DHB), 80 % ACN, 6 % TFA.
4. Enrichment buffer: 80 % ACN, 12 % TFA.
5. Wash buffer A: 10 % ACN, 6 % TFA.
6. Wash buffer B: 40 % ACN, 6 % TFA.
7. Wash buffer C: 60 % ACN, 6 % TFA.
8. Elution buffer A: 5 % ammonium hydroxide solution.
9. Elution buffer B: 10 % ammonium hydroxide solution, 25 % ACN.
10. 100 % Methanol (MeOH).

2.5 Liquid Chromatography and Electrospray Tandem Mass Spectrometry

1. MS buffer A: 0.1 % formic acid (FA).
2. MS buffer B: 0.1 % FA, 80 % ACN.
3. MS buffer C: 1 % TFA, 5 % ACN.
4. 50 cm 75 µm ID fused silica column packed in-house with 1.9 µm reversed-phase C_{18} porous silica beads with 100 Å pore size (Dr. Maisch HPLC GmbH).
5. EASY-nLC 1000 high pressure system (Thermo Fisher Scientific).
6. Q Exactive Plus or Q Exactive HF mass spectrometer (Thermo Fisher Scientific).
7. Data analysis software such as MaxQuant software suite (www.maxquant.org).

3 Methods

3.1 Cell Lysis

Example with adherent HeLa cells grown in 150 mm dishes with a surface area of 17,671 mm^2 (*see* **Note 4**).

1. Heat lysis buffer to 99 °C.
2. Wash cells twice with ice cold PBS and completely aspirate remaining liquid.
3. Add approx. 800 µl of heated lysis buffer and use cell scraper to harvest cells (*see* **Note 5**). Use a pipette to transfer lysate into a centrifuge tube (*see* **Note 6**).
4. Boil in a heat block (99 °C) for 10 min.

5. Sonicate with micro tip probe for 2 min with pulses of 1 s on and 1 s off at an amplitude of 50 %.
6. Measure concentration with Bradford, BCA, or other protein assay to estimate total protein amount (*see* **Note 7**).

   STORAGE OPTION! After this you can store your lysate in the freezer (−20 °C) for a few weeks.

3.2 Protein Digestion

1. Digest lysate with LysC:protein ratio 1:100 (w/w) for 30 min up to 4 h.
2. Dilute to a final concentration of maximum 2 M GuHCl (add minimum two times volume) with 25 mM Tris–HCl pH 8.5 and digest with trypsin 1:100 for 30 min to overnight (*see* **Note 8**).
3. Acidify with TFA to a final concentration of 1 % TFA (pH 1–2).
4. Spin down at max speed for 5 min to pellet insoluble material and keep supernatant (*see* **Note 9**).

3.3 Peptide Desalting

Example with Sep-Pak C_{18} classic cartridge. Adjust the volumes according to the product you are using.

1. Wash Sep-Pak with 3 ml wash buffer I followed by two times 3 ml with wash buffer II using gravity flow.
2. Load sample and after that wash twice with 3 ml wash buffer II.

   STORAGE OPTION! You can keep samples on Sep-Pak in the fridge for 1 week without detectable losses.
3. Elute from Sep-Pak with 3 ml elution buffer I followed by 3 ml elution buffer II.
4. SpeedVac to half volume and measure concentration with Nanodrop A280.

   STORAGE OPTION! Samples can be kept in the freezer for a few months.

3.4 Phosphopeptide Enrichment

1. Prepare in advance C_8 single-layered StageTips (one per sample per incubation) and double-layered C_{18} StageTips (one per sample per incubation) (*see* **Note 3**).
2. Prepare TiO_2 bead slurry to a 1:2 sample:bead ratio (*see* **Notes 10** and **11**).
3. Double your sample volume with enrichment buffer.
4. Add bead slurry and incubate with rotation for 15 min (*see* **Note 12**).
5. Spin down and collect supernatant for second incubation or discard. Leave 20–50 µl of liquid with the beads.

5a. For second and/or third incubation: repeat from **step 4** with the supernatant from **step 5** (*see* **Note 13**).

6. Transfer the bead-peptide suspension to C_8 tips and wash with 50–100 µl wash buffers A, B, and C, respectively, by centrifugation, approximately 500 × *g* for 2 min.
7. Elute phosphopeptides into a 96-well plate or a micro tube with 20 µl of elution buffer A and 20 µl of elution buffer B.
8. SpeedVac until ~5 µl is left.
9. Activate C_{18} StageTips with 50 µl MeOH, followed by 50 µl MS buffer B and 2 × 50 µl MS buffer A.
10. Add 20 µl of MS buffer C to the concentrated sample and load on activated C_{18} StageTips.
11. Wash with 50 µl wash buffer II and elute with 15 µl elution buffer I followed by 15 µl elution buffer II. SpeedVac until <5 µl and add up to 8 µl in total with MS buffer C.

3.5 LC-MS Setup

Example with EASY-nLC 1000 and Q Exactive Plus.

Column was packed in-house using 50 cm long fused silica with 75 µm inner diameter and packed with 1.9 µm C_{18} beads (*see* **Note 14**).

1. Load 5 µl of sample on the column and separate with a MS buffer A and B gradient [% B] going from 5 % to 25 % over 110 min, 25 % to 40 % over 25 min and then a washing step going from 40 % to 80 % over 5 min, staying at 80 % for 5 min and then down to 5 % over 5 min. The flow rate is constant at 250 nl/min.
2. Run the Q Exactive mass spectrometer in the data dependent mode using a top 10 method. Resolution for full scans 70,000, target value 3,000,000, maximum injection time of 20 ms, and scan range from 300 *m/z* to 1750 *m/z*. Resolution for MS-MS scans 35,000, target value 100,000 maximum injection time of 108 ms with normalized collision energy at 28, and a dynamic exclusion of 30 s.

3.6 Data Analysis

1. Analyze the raw files in the free software Max Quant [9] with the Andromeda search algorithm against an appropriate database (e.g., Uniprot or Swissprot human database) for protein identification. Use default settings and add phosphorylation of serine, threonine and tyrosine as variable modification. When running multiple similar phospho raw file use the "match between runs" option in the Global parameter tab for better coverage.
2. Results can be read out from the modificationSpecificPeptides and Phospho (STY) Sites tab-delimited txt output files generated by Max Quant (*see* also Chapters 21 and 22).

4 Notes

1. You can store 6 M GuHCl in the fridge for weeks but TCEP and CAA should be added immediately before use.
2. We use Sep-Pak C_{18} 1 cc cartridge from Waters for up to 1 mg and Sep-Pak C_{18} Classic cartridge for up to 10 mg of protein.
3. Place one or two disks of desired material in a 200 µl pipette tip. A video on how to make StageTips is provided here: http://www.biochem.mpg.de/226863/Tutorials *Note to video*. No special centrifuge is required. Adapters for microcentrifuge tubes can be acquired from Sonation upon special request.
4. If you work with suspension cells skip Subheading 3.1, **steps 2** and **3**, and harvest and wash cell pellet before addition of the heated lysis buffer and proceed from here directly to Subheading 3.1, **step 4**.
5. Adjust volume of lysis buffer to dish or flask size while trying to keep it as low as possible to obtain the most concentrated sample. For suspension cells like HeLa S3 you can use approximately 100 µl per 1,000,000 cells.
6. Be aware that the tube must be compatible with the heat block.
7. Make sure that the method for determining protein concentration is compatible with GuHCl, for example Bradford or BCA.
8. The lower the concentration of GuHCl is the more active trypsin is, so if you are not volume limited dilute up to 10×.
9. Occasionally there is a precipitate appearing after acidification. This will typically not affect the number of identifiable phosphopeptides, but only affect the total protein yield.
10. Example: Use 20 µl bead slurry per 1 mg sample, thus the bead slurry has to be 2 mg/20 µl. This has been tested with up to 20 mg beads per sample. However there is a maximum amount of beads and protein that can be loaded in each StageTip in order to be able to pass liquid through for elution.
11. You can specifically enrich for multiply phosphorylated peptides by adjusting the sample to bead ratio [10]. Usage of scarce amounts of TiO_2 beads typically results in increased number of identified of multiply phosphorylated peptides at the cost of low number of singly phosphorylated peptides. For HeLa lysates, Li et al. (2009) reported an optimal peptide-to-beads ratio of 1:2–1:8 (mass/mass) for obtaining the highest enrichment selectivity and maximum phosphopeptide identifications [10].
12. The beads settle very fast, be sure to have evenly distributed slurry by shaking or tapping the vial before every pipetting event.

13. You can dilute the remaining bead slurry 1:1 with 80 % ACN and 6 % TFA for exceeding incubations.
14. For single shot experiments deeper coverage is obtained with longer columns and gradients. Shorter gradients and 15 cm columns can be used when samples are fractionated.

Acknowledgements

The authors would like to thank members of the Proteomics Program at the Novo Nordisk Foundation Center for Protein Research (CPR) for critical input on the protocol. Work at CPR is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001).

References

1. Macek B, Mann M, Olsen JV (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol 49:199–221. doi:10.1146/annurev.pharmtox.011008.145606
2. Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20(6):261–268
3. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods 11(3):319–324. doi:10.1038/nmeth.2834
4. Greene RF Jr, Pace CN (1974) Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. J Biol Chem 249(17):5388–5393
5. Lippincott J, Apostol I (1999) Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. Anal Biochem 267(1):57–64. doi:10.1006/abio.1998.2970
6. Poulsen JW, Madsen CT, Young C, Poulsen FM, Nielsen ML (2013) Using guanidine-hydrochloride for fast and efficient protein digestion and single-step affinity-purification mass spectrometry. J Proteome Res 12(2):1020–1030. doi:10.1021/pr300883y
7. Francavilla C, Hekmat O, Blagoev B, Olsen JV (2014) SILAC-Based Temporal Phosphoproteomics. Methods Mol Biol 1188:125–148. doi:10.1007/978-1-4939-1142-4_10
8. Kelstrup CD, Young C, Lavallee R, Nielsen ML, Olsen JV (2012) Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. J Proteome Res 11(6):3487–3497. doi:10.1021/pr3000249
9. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372. doi:10.1038/nbt.1511
10. Li QR, Ning ZB, Tang JS, Nie S, Zeng R (2009) Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. J Proteome Res 8(11):5375–5381. doi:10.1021/pr900659n

Part V

High-Throughput Studies of the Phosphoproteome

Chapter 18

Identification of Direct Kinase Substrates via Kinase Assay-Linked Phosphoproteomics

Liang Xue, Justine V. Arrington, and W. Andy Tao

Abstract

Protein phosphorylation plays an essential role in the regulation of various cellular functions. Dysregulation of phosphorylation is implicated in the pathogenesis of certain cancers, diabetes, cardiovascular diseases, and central nervous system disorders. As a result, protein kinases have become potential drug targets for treating a wide variety of diseases. Identification of kinase substrates is vital not only for dissecting signaling pathways, but also for understanding disease pathologies and identifying novel therapeutic targets. However, identification of *bona fide* kinase substrates has remained challenging, necessitating the development of new methods and techniques. The kinase assay linked phosphoproteomics (KALIP) approach integrates in vitro kinase assays with global phosphoproteomics experiments to identify the direct substrates of protein kinases. This strategy has demonstrated outstanding sensitivity and a low false-positive rate for kinase substrate screening.

Key words KALIP, pepKALIP, proKALIP, Kinase substrate, Kinase assay, Phosphoproteomics, Phosphopeptide enrichment

1 Introduction

Protein phosphorylation is an essential posttranslational modification that regulates almost every aspect of biological functions, including signal transduction, protein-protein interactions, and protein translocation [1, 2]. Deregulation of phosphorylation dynamics within the cell often leads to the development of diseases such as cancer, diabetes, immune disorders, and central nervous system pathologies [3]. Therefore, understanding how this modification functions and how it is disturbed in disease states requires phosphorylation analysis, which includes the identification of phosphorylated proteins and corresponding sites of phosphorylation as well as the quantitative measurement of phosphorylation changes [4, 5]. However, detection and quantification of phosphorylation is often challenging because of the low stoichiometry of phosphorylation on proteins [6, 7].

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_18,

Recent advances in mass spectrometry (MS), including improvements in accuracy, sensitivity, and throughput, have allowed it to become an important tool for complex sample analysis. As MS-based proteomics has become popular for large-scale and unbiased analyses [8], MS-based phosphoproteomics has become the method of choice for the examination of global phosphorylation trends [9]. Although current liquid chromatography-mass spectrometry (LC-MS)-based phosphoproteomics has enabled the identification and quantification of thousands of phosphorylation events [10], large-scale proteomics does not typically reveal precise relationships between protein kinases and their direct substrates [11, 12]. In other words, the specific connections between the majority of protein kinases and identified phosphorylation sites have yet to be elucidated. Thus, mapping kinase-substrate relationships is a critical step for understanding essential signaling networks and identifying pharmaceutical targets for drug discovery [11].

Here, we describe an integrated strategy termed kinase assay-linked phosphoproteomics (KALIP) for identifying the direct substrates and substrate specificity of protein kinases with high sensitivity and confidence. The in vitro kinase reaction is carried out using either peptides (pepKALIP) or proteins (proKALIP) that are derived directly from cellular proteins and then dephosphorylated to serve as substrate candidates. The resulting newly phosphorylated peptides are then isolated and identified by mass spectrometry. At the same time, global phosphoproteomics experiments with cells in which the kinase is either active or inhibited reveal the kinase-dependent phosphoproteome. The overlap between the direct substrates identified by in vitro kinase assays and the kinase-dependent endogenous substrates represents *bona fide* substrates of the target kinase.

In the peptide-level KALIP (pepKALIP) procedure (left branch of Fig. 1), the cell lysate is first enzymatically digested to generate a peptide pool. When a tyrosine kinase is being investigated, we recommend pre-enriching tyrosine phosphopeptides before the following biochemical reactions. An alkaline phosphatase is used to dephosphorylate endogenous phosphopeptides so that the background phosphorylation level is maximally reduced in the cell lysate. After deactivation of the phosphatase by heat shock, the target kinase is added to phosphorylate its substrates. A common phosphoproteomics workflow including peptide desalting, phosphopeptide enrichment and LC-MS analysis is sequentially applied. The data that is produced contains in vitro direct peptide substrates of the kinase.

In the protein-level KALIP (proKALIP) procedure (*see* middle branch of Fig. 1), instead of enzymatically digesting the cell lysate prior to biochemical reactions, the protein extracts go through dephosphorylation and rephosphorylation by an in vitro kinase

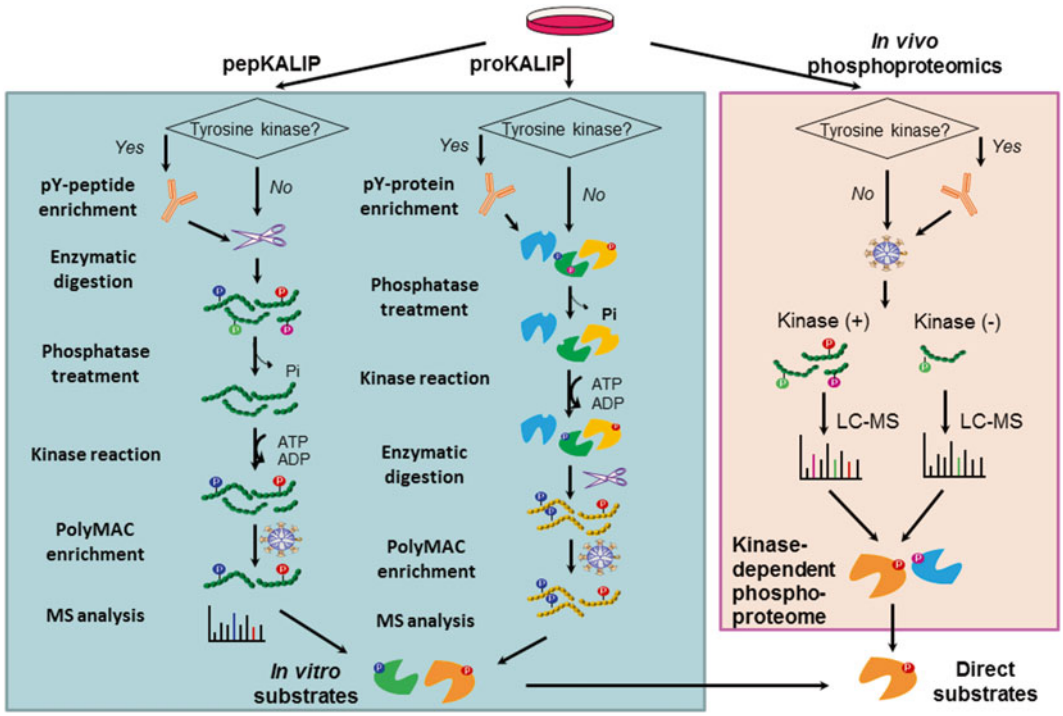


Fig. 1 KALIP strategy: Workflow for kinase substrate identification through the integration of in vitro kinase reactions and in vivo phosphoproteomics. In the in vitro kinase reaction (*blue panel*), the peptides (*left branch*) or proteins (*middle branch*) from the cell lysate are dephosphorylated before the kinase assay. Phosphopeptides are enriched and analyzed by mass spectrometry for sequencing and site identification. Though in vivo phosphoproteomics (*pink panel*), kinase-dependent phosphorylation events are identified by comparing wild-type and kinase-inhibited cells. *Bona fide* direct substrates are the overlapping phosphoproteins present in both in vitro and in vivo datasets

assay before trypsin digestion. We also recommend an extra pre-enrichment step for tyrosine kinases. The dephosphorylation step is the same as in pepKALIP procedure discussed previously. However, the in vitro kinase assay requires addition of a generic irreversible kinase inhibitor to block endogenous kinases present in the total cell lysate. It is imperative that the excess kinase inhibitor is removed before adding the kinase of interest. Once the in vitro kinase assay is complete, the protein extracts with the exogenous kinase and phosphatase are digested by trypsin, followed by peptide desalting, phosphopeptide enrichment, and LC-MS analysis. The resulting data contains in vitro direct protein substrates of the kinase.

Both pepKALIP and proKALIP require datasets from a phosphoproteomics study (right branch of Fig. 1). Ideally, the phosphoproteome from cells with kinase perturbation (e.g., inhibition or genetic knockout of the kinase, activation or overexpression of

the kinase) is compared to that of a control. The phosphoproteomic experiments generate a set of proteins whose phosphorylation is dependent on the kinase of interest. Lastly, the overlap between the direct substrates identified from an in vitro kinase reaction and the in vivo kinase-dependent phosphorylation identified from global phosphoproteomics represents *bona fide* direct substrates of the target kinase.

2 Materials

Ultrapure water and analytical grade reagents should be used to make all solutions. Unless otherwise noted, the solutions here were prepared using reagents from Sigma-Aldrich and nanopure water that was made by purifying deionized water.

2.1 Cell Lysis

1. Frozen cell pellet to provide at least 500 μg of protein. Most mammalian cell lines (e.g., DG-75 or HeLa) will work.
2. Lysis buffer-ST for Ser/Thr kinase: 50 mM Tris–HCl, pH 7.5; 150 mM NaCl, 1 % NP-40.
3. Lysis buffer-Y for Tyr kinase: 50 mM Tris–HCl, pH 7.5; 150 mM NaCl, 1 % NP-40, 1 mM sodium orthovanadate.
4. Tabletop centrifuge.

2.2 Protein Digestion

1. BCA assay reagents.
2. 8 M urea, 50 mM Tris–HCl, pH 8.0.
3. RapiGest surfactant: RapiGest is freshly prepared and used immediately (*see* **Note 1**).
4. 1 M trimethylammonium bicarbonate (TMAB) buffer.
5. 200 mM dithiothreitol (DTT) (dissolved in water and stored in single-use aliquots at −20 °C).
6. 300 mM iodoacetamide (freshly prepared in water and used immediately).
7. Proteomics-grade trypsin dissolved in 50 mM acetic acid (stored at −20 °C until use) or directly dissolved in the lysate.
8. 1 M HCl (stored at room temperature).
9. Sep-Pak C18 Solid-Phase Extraction (SPE) Cartridges.
10. 80 % acetonitrile, 0.1 % trifluoroacetic acid (TFA).
11. Thermal shaker.
12. Vacuum centrifuge.

2.3 Enrichment of Phosphotyrosine-Containing Peptides

1. 1 M Tris–HCl pH 8.0.
2. Anti-phosphotyrosine antibody (clones PT66 and PY20), immobilized on agarose beads in a 1:1 slurry (Sigma-Aldrich).

3. Elution solutions: 0.1 % trifluoroacetic acid (TFA); 0.1 % TFA, 50 % acetonitrile; 100 mM glycine (pH decreased to 2.5 with TFA).
4. Sample rotator/sample spin wheel.

2.4 Enrichment of Phosphotyrosine-Containing Proteins

1. 1 M Tris–HCl, pH 8.0.
2. Anti-phosphotyrosine antibody (clone PT66), immobilized on agarose beads in a 1:1 slurry (Sigma-Aldrich).
3. Elution solution: 100 mM triethylamine.
4. Sample rotator/sample spin wheel.

2.5 In Vitro *Kinase Reaction (See* Note 2*)*

1. Dephosphorylation buffer: Thermosensitive phosphatase in 1× phosphatase buffer such as the rAPid Alkaline Phosphatase kit from Roche.
2. Kinase reaction solution: Purified kinase, $MgCl_2$, 1 mM ATP, and other supplements required for kinase activity in 50 mM Tris–HCl buffer.

2.6 Phosphopeptide Enrichment Using PolyMAC-Ti

1. Polymer-based Metal-ion Affinity Capture (PolyMAC) Phosphopeptide Enrichment Kit (Tymora Analytical, West Lafayette, IN).

2.7 Mass Spectrometry and Data Analysis

1. 0.1 % formic acid in water.
2. LC-MS buffers: Buffer A: 0.1 % formic acid in water. Buffer B: 0.1 % formic acid in acetonitrile.
3. Integrated electrospray emitter tip: Prepared by packing 30 cm of 75 μm inner diameter fused-silica capillary with 3 μm ProntoSIL C18-AQ resin (Bischoff Chromatography, Leonberg, Germany).
4. LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher) or equivalent.
5. Laser puller (Model P-2000; Sutter Instrument Co.).

3 Methods

3.1 Cell Lysis

1. Lyse the cells by resuspending them in 1 mL of the lysis solution (*see* **Notes 3** and **4**).
2. Centrifuge the lysed cells at 16,000 × *g* and collect the supernatant. This step eliminates the insoluble fractions of the cell.

3.2 Protein Digestion *(See* Note 5*)*

1. Determine the protein concentration of the lysates by BCA assay (following the manufacturer's protocol).
2. Resuspend the dried RapiGest surfactant aliquot with 1 M trimethylammonium bicarbonate (TMAB) solution and add to

the lysate to bring the final concentration of RapiGest to 0.1 % (w/v) and TMAB to 50 mM.

3. Add DTT solution to a final concentration of 5 mM and incubate for 30 min at 50 °C to reduce the disulfide bonds.
4. Cool samples to room temperature and alkylate the free sulfide groups by incubating the samples in 15 mM of iodoacetamide for 1 h at room temperature in the dark.
5. To digest the proteins, add proteomics-grade trypsin to the samples at a 1:100 ratio and incubate for 10–16 h at 37 °C (*see* **Note 6**).
6. After digestion, add sufficient 1 M HCl to the samples to attain a final concentration of 100–150 mM in order to bring the pH < 3. Incubate at 37 °C for 45 min to precipitate the RapiGest out of solution.
7. Spin down the samples at 16,000 × *g* for 10 min and collect the supernatant, thus removing RapiGest.
8. Desalt the peptides by using a Sep-Pak C18 SPE column following the manufacturer's protocol (*see* **Note 7**).
9. Elute the peptides off the column with 80 % acetonitrile in 0.1 % TFA. Dry the eluates completely using a vacuum centrifuge concentrator and freeze until further use.

3.3 Enrichment of Phosphotyrosine-Containing Peptides

1. Phosphotyrosine-containing peptides can be enriched by anti-phosphotyrosine immunoprecipitation before *in vitro* kinase assays (*see* **Note 8**).
2. After the RapiGest removal (*see* **step 6**), add sufficient 1 M Tris–HCl (pH 8.0) to increase the pH to 7.2–7.4. Add 40 μL PT-66 clone anti-pTyr-agarose beads slurry for each 1 mg of protein in the sample, and incubate 12–16 h at 4 °C by end-over-end rotation (*see* **Note 9**).
3. Spin down the beads at 400 × *g* for 30 s and remove the supernatant.
4. Wash the beads by adding 500 μL of lysis buffer and incubate 10 min at 4 °C by end-over-end rotation. Spin down the beads at 400 × *g* for 30 s, remove the supernatant, and repeat.
5. Rinse the beads once with water, remove the remaining solution, and add the first elution buffer—100 μL of 0.1 % TFA. Incubate for 10 min by shaking at room temperature.
6. Spin down the beads 400 × *g* for 30 s, and collect the supernatant in a low-binding microfuge tube. Repeat the elution process two more times with 100 μL of 0.1 % TFA and another two times with 100 μL of 0.1 % TFA and 50 % acetonitrile. Collect the eluents into the same tube. Add 50 μL of the final elution solution, 100 mM glycine (pH 2.5), and incubate the

beads by shaking for 30 min. Spin down, collect the supernatant into the eluent tube, and repeat one more time. Dry the eluents completely using a vacuum centrifuge concentrator. The phosphotyrosine containing-peptides are ready for the kinase reaction (*see* Subheading 3.5) or for PolyMAC-Ti enrichment (*see* Subheading 3.6).

3.4 Enrichment of Phosphotyrosine-Containing Proteins

1. Phosphotyrosine-containing proteins can be enriched by anti-phosphotyrosine immunoprecipitation before *in vitro* kinase assay (*see* **Note 8**).
2. After cell lysis, (*see* Subheading 3.1) and before protein digestion, add 40 μL PT-66 clone anti-pTyr-agarose beads slurry for each 1 mg of protein in the sample and incubate for 12–16 h at 4 °C by end-over-end rotation.
3. Spin down the beads at 400 × *g* for 30 s, and remove the supernatant.
4. Wash the beads by adding 500 μL of lysis buffer and incubate 10 min at 4 °C by end-over-end rotation. Spin down the beads at 400 × *g* for 30 s, remove the supernatant, and repeat.
5. Rinse the beads once with water, remove the remaining solution, and add the first elution buffer: 100 μL of 100 mM triethylamine (TEA). Incubate for 30 min under vigorous shaking at 4 °C.
6. Spin down the beads 400 × *g* for 30 s, and collect the supernatant in a low-binding microfuge tube. Repeat the elution process one more time. Collect the eluents into the same tube.
7. Decrease the eluent volume down to 20 μL using a vacuum centrifuge concentrator. Add sufficient 1 M Tris–HCl (pH 8.0) to increase the pH to 7.2–7.4. The phosphotyrosine protein-containing solution is ready for the subsequent kinase reaction (*see* Subheading 3.5).

3.5 Kinase Reaction (*See* Note 10)

1. Resuspend samples of phosphopeptides in 200 μL of phosphatase buffer.
2. Add two units of thermosensitive phosphatase and incubate at 37 °C for 1–2 h. Deactivate the phosphatase by heating at 75 °C for 5 min.
3. Incubate samples in buffer containing the kinase of interest, 5 mM $MgCl_2$, and 1 mM ATP at 30 °C for 1 h (*see* **Note 2**). Reactions can be quenched by the addition of 1 % TFA to a pH < 3.
4. If performing pepKALIP, the samples are ready for phosphopeptide enrichment by PolyMAC-Ti (*see* Subheading 3.6). If proKALIP is being done, the samples are ready for protein digestion (*see* Subheading 3.2) followed by PolyMAC-Ti enrichment.

3.6 Phosphopeptide Enrichment Using PolyMAC-Ti Kit

1. The peptide samples are first desalted using Sep-Pak C18 columns and dried (*see* Subheading 3.2, **step 8**).
2. Mix the capturing beads well, carefully transfer 50 μL of the slurry to an Eppendorf tube in a magnetic rack, and remove the storage solution.
3. Wash the beads twice with 200 μL of nanopure water.
4. Add 100 μL of the PolyMAC-Ti loading buffer into the peptide sample and resuspend well by vortexing and shaking.
5. Add 10 μL of PolyMAC-Ti reagent, vortex for 10 s, and shake the solution for 5 min.
6. Add 200 μL of the capture buffer to increase the pH above 6.3. Vortex briefly.
7. Pipette the mixture up and down a few times, and transfer the whole solution to the capturing beads in the tube. Shake for 10 min.
8. Use magnetic rack and discard the flow-through.
9. Incubate the beads by shaking with 200 μL of PolyMAC loading buffer for 5 min and discard the supernatant.
10. Repeat **step 9**, Subheading 3.6 twice with the washing buffer and once with nanopure water.
11. Incubate the beads twice by shaking with 100 μL of elution buffer for 5 min each time.
12. Collect the eluents in the same low-binding tube. Dry the eluent completely using a vacuum centrifuge concentrator.

3.7 Mass Spectrometry and Data Analysis

1. Complete LC-MS analysis of peptide samples. As an example, we employ an Easy-nLC 1000 system coupled to an LTQ-Orbitrap Velos mass spectrometer. The reverse-phase C18 separation is performed using an in-house capillary column packed with 3 μm C18 beads. The mobile phase buffer consists of 0.1 % formic acid in ultrapure water with an eluting buffer of 0.1 % formic acid (buffer A) in 80 % acetonitrile (buffer B) run over a 90-min linear gradient at a flow rate of 300 nL/min. The electrospray ionization emitter tip is generated on the prepacked column with a laser puller. The mass spectrometer is operated in data-dependent mode in which a full-scan MS (from m/z 300–1700 with the resolution of 60,000 at m/z 400) is followed by 20 MS/MS scans of the most abundant ions. Ions with charge state of +1 are excluded. The mass exclusion time is 30 s.
2. Complete a database search to identify phosphorylation sites by the kinase of interest. As an example, our MS data are searched against the desired proteome database using the SEQUEST algorithm with a static modification of +57.021 Da

on Cys and a variable modification of +79.996 Da on Ser, Thr, and Tyr. The digestion enzyme is specified within the search, and the false discovery rate (FDR) is set to 1 % for each analysis. Phosphorylation site localization is further determined by site localization programs.

3. The data containing substrate information can be analyzed with various bioinformatics tools such as Motif-X [12] and Ingenuity Pathway Analysis (Ingenuity Systems), depending on the desired information.

4 Notes

1. Because RapiGest is not stable in water and cannot withstand multiple freeze-thaw cycles, it is advisable to dry freeze it in 1 mg working aliquots.
2. For the protocol of kinase assay, including buffer recipe, enzyme/substrate amount, and reaction time, please refer to literature about the kinase of interest or do some preliminary tests to determine the optimal conditions. For example, a kinase assay with the spleen tyrosine kinase (Syk) can be done with pre-enriched phosphotyrosine containing peptides in 50 mM Tris–HCl (pH 7.2–7.4) at 30 °C with shaking for 30 min. A reaction with 300 ng of Syk kinase also requires 5 mM $MgCl_2$ and 1 mM ATP and can be quenched by decreasing the pH < 3 with 1 % TFA [13]. The same reaction conditions can be used for a protein level kinase assay with Syk, although the reaction can be quenched with 8 M urea in 5 mM DTT to produce conditions that are more amenable to direct protein digestion [14].
3. For investigation of tyrosine kinases, we recommend treating the cells with pervanadate before harvest to increase the overall tyrosine phosphorylation level. This step increases the sensitivity and yield of the following phosphotyrosine peptide/protein enrichment, so the maximum number of candidate substrates can be isolated before the subsequent biochemical assays.
4. It is best to use the lysate immediately because the level of phosphorylation decreases even if the sample is stored at −80 °C
5. Before digestion, ensure that pH of the sample is around 8; if lower, add more 1 M TMAB to increase the pH.
6. The choice of digestion enzyme depends on the specificity of target kinase. Ideally, the cleavage site should not be included in the theoretical motif of the kinase.
7. Briefly, our desalting procedure consists of wetting the resin with one column volume of methanol followed by one column volume of 0.1 % TFA in 80 % acetonitrile. The organic solvent

is washed out with at least three column volumes of 0.1 % TFA. We load our acidified peptide samples and then wash the column with at least three column volumes of 0.1 % TFA. More washes may be needed for samples with high salt concentrations. The peptides are then eluted with one column volume of 0.1 % TFA in 80 % acetonitrile and dried to completion with a vacuum centrifuge concentrator.

8. Several benefits are gained by pre-enriching of phosphotyrosine peptides or proteins. First, the collection of substrates is derived from formerly tyrosine phosphorylated proteins. This extra step efficiently reduces the basal phosphorylation which may interfere with the kinase reaction and result in a high rate of false positives. Second, application of the effective phosphatase inhibitor pervanadate elevates phosphorylation levels, allowing isolation of a large number of phosphoproteins and increasing the overall sensitivity of the strategy. Third, the generic alkaline phosphatase has higher efficacy with a purified phosphotyrosine proteome than with crude cell lysates. Pre-enrichment helps phosphotyrosine peptides or proteins to be specifically dephosphorylated instead of using large amounts of phosphatase to dephosphorylate all of the phosphopeptides or proteins in the cell extracts. This helps to eliminate the sizeable serine and threonine phosphoproteome, which would not be target of tyrosine kinases. Fourth, the enrichment yields the peptides or proteins that have been actually phosphorylated within the intact cell, excluding peptides or proteins that may be exclusively phosphorylated *in vitro*.
9. To improve anti-phosphotyrosine immunoprecipitation, a combination of different antibody clones can be used simultaneously (e.g., PT66 together with pY20) in order to ensure a more complete coverage of the phosphotyrosine-containing peptides.
10. In the case of proKALIP the *in vitro* kinase assay requires addition of a generic irreversible kinase inhibitor to block endogenous kinases present in the total cell lysate. One way we inhibit endogenous kinases is to treat dephosphorylated proteins with 1 mM 5′-(4-fluorosulfonylbenzoyl)adenosine (FSBA) with 10 % DMSO in Tris–HCl, pH 7.5 at 30 °C for 1 h. Excess FSBA is removed by centrifugal filtration units (30 kDa MW cutoff).

Acknowledgments

The authors gratefully acknowledge that this work has been funded in part by an NSF CAREER award CHE-0645020 (WAT), and by National Institutes of Health grant GM088317 (WAT).

References

1. Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell 116:191–203
2. Hunter T (2000) Signaling—2000 and beyond. Cell 100:113–127
3. Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. Nature 411:355–365
4. Cohen P (2001) The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. Eur J Biochem 268:5001–5010
5. Manning BD, Cantley LC (2002) Hitting the target: emerging technologies in the search for kinase substrates. Sci STKE 162:pe49
6. Wu RH, Haas W, Dephoure N, Huttlin EL, Zhai B, Sowa ME, Gygi SP (2011) A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat Methods 8:677–U111
7. Jin LL, Tong JF, Prakash A, Peterman SM, St-Germain JR, Taylor P, Trudel S, Moran MF (2010) Measurement of Protein Phosphorylation Stoichiometry by Selected Reaction Monitoring Mass Spectrometry. J Proteome Res 9:2752–2761
8. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207
9. Mumby M, Brekken D (2005) Phosphoproteomics: new insights into cellular signaling. Genome Biol 6:230
10. Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14:35–48
11. Cohen P (2002) Protein kinases–the major drug targets of the twenty-first century? Nat Rev Drug Discov 1:309–315
12. Schwartz D, Gygi SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol 23:1391–1398
13. Xue L, Wang W, Iliuk A, Hu L, Galan JA, Yu S, Hans M, Geahlen RL, Tao WA (2012) Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. Proc Natl Acad Sci U S A 109:5615–5620
14. Xue L, Geahlen RL, Tao WA (2013) Identification of direct tyrosine kinase substrates based on protein kinase assay-linked phosphoproteomics. Mol Cell Proteomics 12:2969–2980

Chapter 19

Phosphoprotein Detection by High-Throughput Flow Cytometry

Johannes Landskron and Kjetil Taskén

Abstract

Phospho flow cytometry is a powerful technique for the detection of protein phosphorylation events that, like Western blotting, relies on phospho-epitope-specific antibodies. In contrast to the latter, however, multidimensional and directly quantifiable data is obtained at the single-cell level allowing separate analysis of small cell populations in complex cellular mixtures. Furthermore, up to 30 phospho-specific antibodies or antibodies identifying other posttranslational modifications in combination with cell surface markers can be analyzed in a single experiment. Utilizing a technique called fluorescent cell barcoding that enables combination of up to 64 samples into one tube for multiplex analysis and later data deconvolution, phospho flow cytometry is turned into a medium- to high-throughput technology.

Key words Flow cytometry, Phospho flow, Phospho-specific antibody, Fluorescent cell barcoding, Signaling

1 Introduction

Proteomics approaches to analyze protein phosphorylation and other posttranslational modifications of proteins are principally unbiased and result in large amounts of data. However, key findings from such analyses have to be validated using independent techniques. Methods to validate findings in proteomics data sets are typically biased approaches involving specific detection and quantification of proteins of interest by antibodies. Western blotting appears to still be the method of choice for many such purposes. However, especially for phosphoproteomics and signaling network analysis, flow cytometry utilizing phospho-epitope-specific antibodies for intracellular staining [1], called *phospho flow*, is gaining importance [2, 3]. Although phospho flow requires costly equipment and, like Western blotting, is limited to the availability of antibodies, it offers many advantages over the latter. Data is recorded electronically at the single cell level, can be directly quantified and multidimensional analyses are possible. Today's

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_19,

standard instruments offer, depending on the number of lasers and detectors, usually at least 6–8 color channels in addition to forward and side scatter parameters. However, instruments equipped with up to 5 lasers, recording as many as 20 channels are available. These color channels can be used for various approaches. Surface markers can be included to identify numerous cell types, such as different lymphocyte populations in a blood-derived sample. Furthermore, samples from different treatment conditions, like stimulation time courses, concentration-responses and perturbations can be labeled with varying intensities of one or several fluorochromes, combined into a single sample and deconvoluted again during analysis using a protocol called fluorescent cell barcoding (FCB) [4, 5]. Up to 64 samples can be pooled that way, lowering not only the costs by reducing antibody consumption and shortening run-time, but also increasing data quality and comparability by ensuring equal treatment for all FCB populations within one sample. In addition, the combined FCB sample can be split up and stained with panels of up to 30 distinct antibodies, depending on the initial cell number. Taking these three aspects together, several thousand data points can be gathered during a single experiment making phospho flow a true high-throughput methodology for the analysis of protein phosphorylation. In another possible application of the method, the various channels can be used to stain for a panel of phospho-specific antibodies conjugated to different fluorochromes to obtain multidimensional single cell-based signaling data for bioinformatical applications like Bayesian network analysis. In yet another application, the phospho flow technology can be used in cell-based screening of compound libraries to find small molecules that interfere with particular signaling events or to find mechanism of action of small molecules using signal network analysis.

The protocol below describes a relatively small setup of a four time point (0 min, 1 min, 5 min, 15 min) stimulation time course for Jurkat T cells that are treated with a CD3/TCR-stimulating antibody. After fixation with paraformaldehyde, the cells are fluorescently barcoded, combined, permeabilized with methanol to facilitate intracellular staining, labeled with phospho-specific antibodies and analyzed by flow cytometry (*see* Fig. 1). In addition, the procedures for antibody titration and specificity testing, and an adaption of the protocol to adherent cells is described.

2 Materials

2.1 Solutions, Buffers, Cell Culture, and Instrumentation

1. Phosphate-buffered saline (PBS, pH 7.4).
2. Wash solution: PBS 1 % fetal bovine serum and, optionally, 0.09 % sodium azide.

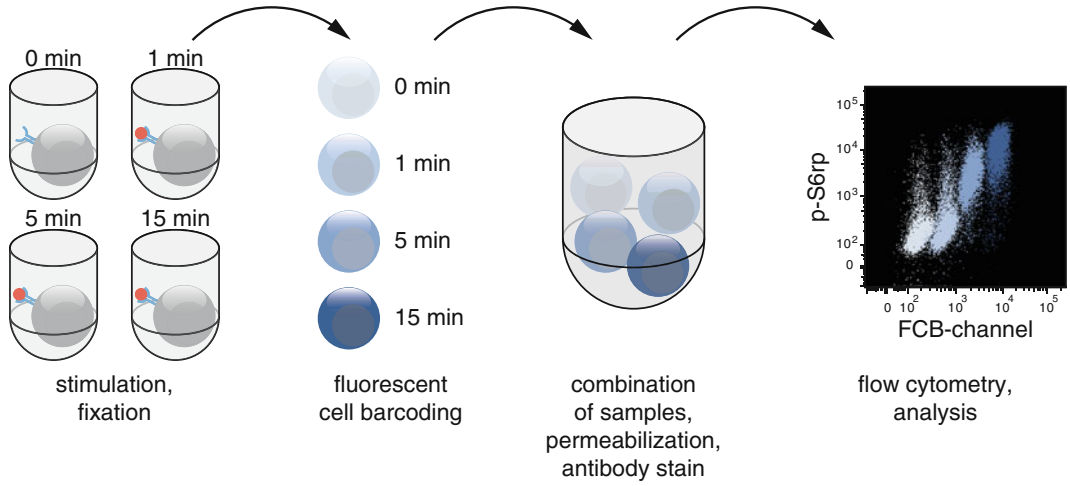


Fig. 1 Overview of the described phospho flow protocol. First, cells are stimulated followed by fixation and fluorescent cell barcoding (FCB). Then, the differentially marked cells are combined into a single tube and permeabilized. After antibody staining, the cells are analyzed by flow cytometry. p-S6rp: S6 ribosomal protein (pS235/236). Note that separation of populations in the FCB channel allows deconvolution of the earlier combined samples

3. Cell culture medium: RPMI 1640, GlutaMAX, 10 % fetal bovine serum, MEM nonessential amino acids solution, 1 mM sodium pyruvate, and penicillin/streptomycin.
4. Fixation solution: 3 % Paraformaldehyde solution or Fix Buffer I (BD Phosflow) (*see* **Note 1**).
5. Permeabilization solution: 100 % methanol or Perm Buffer III (BD Phosflow) (*see* **Notes 2** and **3**).
6. Stimulation agent for Jurkat T cells: OKT3 antibody against CD3ε (ATCC CRL-8001). For primary T-cells: biotinylated anti-CD3ε, biotinylated anti-CD28 (e.g., antibody clone CD28.2, eBioscience), biotinylated anti-CD2 (e.g., antibody clone RPA-2.10, eBioscience), avidin.
7. Cells: Jurkat T cells clone E6-1 (ATCC TIB-152).
8. 96-Well v-bottom plates, 1.5 mL Eppendorf tubes.
9. Centrifuges for 1.5 mL tubes and plates.
10. Temperature controlled water bath.
11. Flow cytometer, equipped with a red (633–670 nm) and a UV or violet (375–413 nm) laser, and optimally a plate loader or high-throughput sampler.

2.2 Fluorescent Dyes and Antibodies

1. Phospho-epitope-specific primary antibodies, optimally conjugated to a fluorescent dye, e.g., Alexa Fluor 647 or unconjugated (*see* **Note 4**).

2. If unlabeled primary antibodies are in the panel: isotype-specific fluorescently labeled secondary antibodies (e.g., Alexa Fluor 647 Goat anti-Rabbit IgG (H+L), Molecular Probes, Life Technologies, 2 mg/mL stock, final dilution 1:8000).
3. Fluorescently labeled isotype control (e.g., Alexa Fluor 647 mouse IgG1, κ isotype control).
4. FCB reagent: DMSO (dimethylsulfoxide), amine-reactive Pacific Blue succinimidyl ester (Molecular Probes, Life Technologies).

2.3 Antibody Titration and Cold Trypsinization

1. PMA and ionomycin: Prepare a 10 mM stock solution from phorbol-12-myristate-13-acetate (PMA) and a 1 mg/ml stock solution from ionomycin calcium salt in DMSO (*see* **Note 5**). Dilute the PMA 1:1200 and the ionomycin 1:10 in complete medium, respectively, to obtain the 100× solutions.
2. Calyculin A, 10 μM solution in DMSO (100× solution) (*see* **Note 5**).
3. Pervanadate: Prepare a 0.1 M sodium orthovanadate Na_3VO_4, aqueous solution. Mix 10 μL with 90 μL of 1 % hydrogen peroxide to obtain the 100× solution (*see* **Note 6**).
4. 2.5 % trypsin (10×).
5. 6-Well plates with cell culture-treated surface.

3 Methods

Samples should be kept on ice unless indicated differently and in darkness when fluorochromes are used.

3.1 Stimulation and Fixation

1. Place 245 μL of a 5×10^7 cells/mL cell suspension in complete cell culture medium in the first well of a 96-well v-bottom plate (*see* **Note 7**) and incubate at 37 °C in a water bath for 15 min (stimulation plate) (*see* **Note 8**).
2. Pipet 10 μL OKT3 (50 μg/mL in cell culture medium) into the next well (*see* **Note 9**).
3. Place a second 96-well v-bottom plate in the water bath with 50 μL fixation solution in the first three consecutive wells (fixation plate) (*see* **Note 10**).
4. Fixation of the (unstimulated) 0 min time point: Mix the cell suspension well by pipetting up and down and transfer 50 μL into the first well with fixative of the fixation plate and mix.
5. Stimulation: Transfer 5 μL of the OKT3 solution into the cell suspension and immediately mix well.
6. Fixation of time points 1 and 5 min: After 1 and 5 min, mix cell suspension well and transfer 50 μL for each time point into the last 2 wells with fixative in the fixation plate and mix.

7. After 13 min: Pipet 50 µL fixation solution in the third well of the stimulation plate (*see* **Note 10**).
8. Fixation of time point 15 min: After 15 min, mix cell suspension well and transfer 50 µL into the fixative in the same plate (*see* **Note 11**).
9. Also after 15 min: Place the fixation plate (time points 0, 1, 5 min) on ice and add 100 µL cold PBS to each well.
10. After 25 min: Add 100 µL cold PBS to the last fixation reaction and transfer the suspension to the next empty well in the fixation plate.
11. Wash step: Centrifuge plate with 700 × *g* for 5 min, remove supernatants, and resuspend cell pellets in 200 µL of PBS each.

3.2 Fluorescent Cell Barcoding

1. Dissolve the Pacific Blue in DMSO to a final concentration of 10 mg/mL. From that, prepare a 200 µL solution in DMSO with a concentration of 100 µg/mL (*see* **Note 12**).
2. Prepare a series of dilutions of the Pacific Blue barcoding reagent starting with the 100 µg/mL stock from above and dilute 1:20 (to 5 µg/mL). Then use the resulting dilutions to next dilute 1:4 (1.25 µg/mL), 1:4 (312.5 ng/mL), and 1:9 (34.72 ng/mL) in DMSO and transfer 5 µL of each into separate wells of a new 96-well v-bottom plate (*see* **Note 13**).
3. Pipet the cell suspensions into the wells with the barcoding reagents. 0 min into the well with 34.72 ng/mL solution, 1 min into 312.5 ng/mL, 5 min into 1.25 µg/mL, and 15 min into 5 µg/mL and directly mix by pipetting up and down several times (*see* **Note 14**).
4. Incubate at room temperature (RT) and in darkness for 30 min. Shaking is not required.
5. Perform two consecutive wash steps (Subheading 3.1, **step 11**), using wash solution instead of PBS.
6. Combine all reaction in a suitable tube (e.g., 1.5 mL micro test tube).

3.3 Permeabilization

1. Centrifuge the combined suspension with 700 × *g* for 5 min, remove the supernatants down to some µL and resuspend the pellet by vortexing (*see* **Note 15**).
2. Add 500 µL cold permeabilization solution, mix, and incubate at −80 °C for 20 min (*see* **Note 16**).
3. Incubate for 10 min on ice and repeat wash step (Subheading 3.1, **step 11**), using wash solution instead of PBS.

3.4 Antibody Staining

1. Resuspend in 50 µL wash solution per antibody stain plus an additional 50–100 µL; for example for 20 stains resolve the pellet in 22 × 50 µL = 1.1 mL wash solution (*see* **Note 17**).

2. Prepare a new plate containing the different phospho-specific antibodies (*see* **Notes 18** and **19**) for each sample in a total volume of 50 µL wash solution. If unlabeled primary antibodies are part of the panel, tubes or a separate plate should be used for these reactions, because a secondary antibody stain will be required.
3. Mix 50 µL at a time of the cell suspension with each of the 50 µL antibody solutions and incubate at RT and in darkness for 30 min. Shaking is not required.
4. Add 100 µL wash solution and repeat Subheading 3.1, **step 11** two times, using wash solution instead of PBS. In case of unlabeled primary antibodies, the final resuspension volume is 50 µL. Stains with fluorescently labeled primary antibodies are ready for flow cytometry after this step.
5. If unlabeled primary antibodies have been used, mix each reaction with 50 µL of a suitable dilution of a fluorescently labeled secondary antibody and incubate at RT and in darkness for 30 min.
6. Add 100 µL wash solution and repeat Subheading 3.1, **step 11** two times, using wash solution. Samples are now ready for flow cytometry.

3.5 Flow Cytometry and Analysis

1. Run the maximal volume of each sample compensated (*see* **Note 20**) on a suitable flow cytometer e.g. BD LSRFortessa equipped with 3 lasers (blue: 488 nm, red: 640 nm and violet: 405 nm) and a high-throughput sampler. Settings: Filters: standard optical setup for the Alexa Fluor 647 (670/30) and the Pacific Blue (450/50) channel, *throughput mode*: Standard, *events to record*: 1,000,000, *sample flow rate*: 1.0 µL/s, *sample volume*: 150 µL, *mixing volume*: 100 µL, *mixing speed*: 200 µL/s, *number of mixes*: 3, *wash volume*: 400 µL. If possible, include the parameters forward and side scatter width and height (FSC-H, FSC-W, SSC-H, and SSC-W) in addition to area (FSC-A, SSC-A) (*see* **Note 21**).
2. Import the data into a flow cytometry analysis software like FlowJo or Cytobank.
3. Gating strategy:
   - Select the *cell* population in the FSC-A/SSC-A density dot plot (*see* Fig. 2a).
   - Display the *cell* population in a FSC-H/FSC-W density plot and draw the gate around the single-cell population *single cells* (*FSC*) (*see* Fig. 2b).
   - Display the *single-cell* (*FSC*) population in a SSC-H/SSC-W density plot and draw the gate around the single-cell population *single cells* (*SSC*) (*see* Fig. 2c).

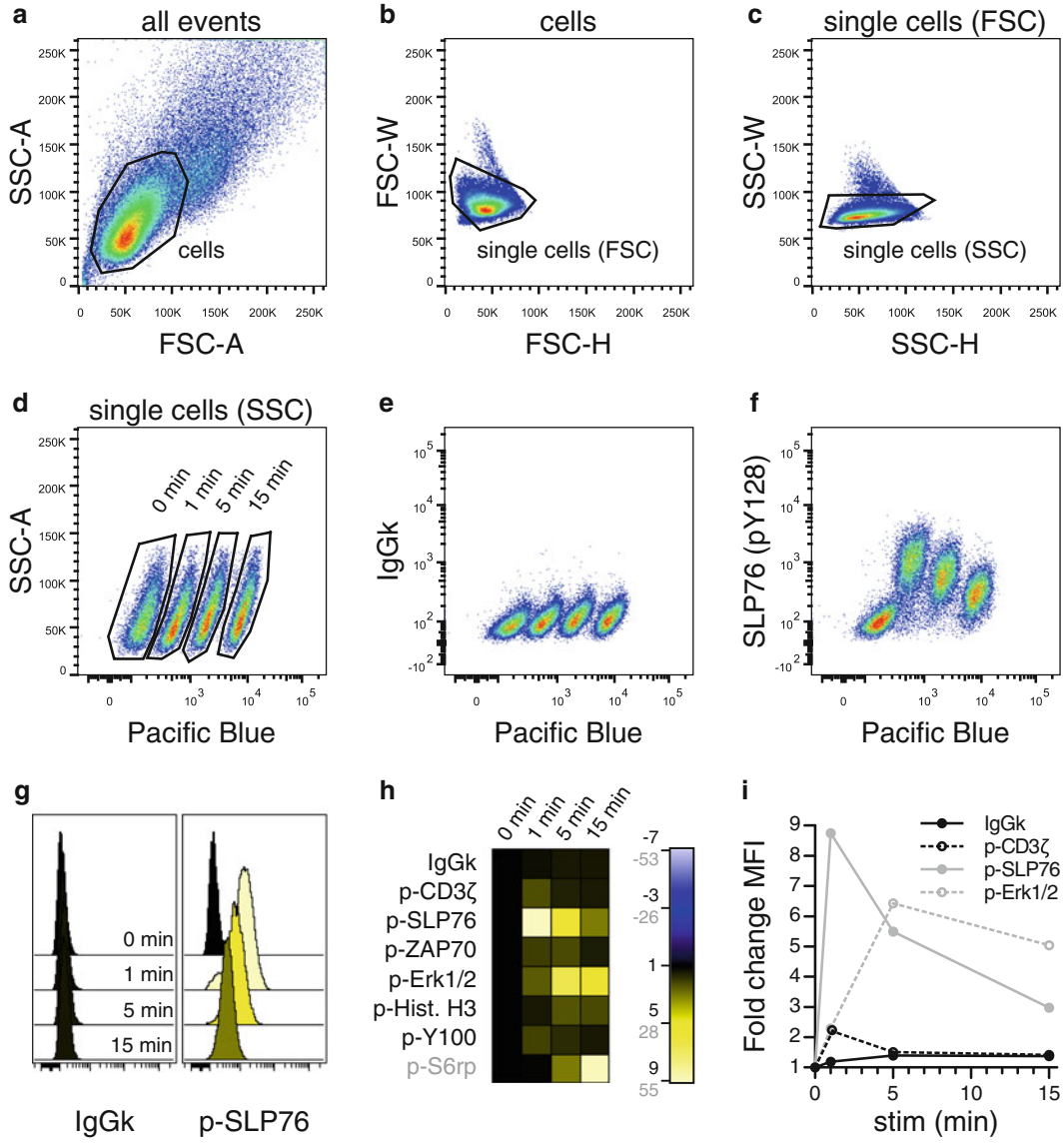


Fig. 2 Analysis of a phospho flow experiment. Gating strategy: Selection of the cell population (**a**), followed by two gating steps to determine the single cells (**b**, **c**) and the barcoding populations based on their Pacific Blue staining intensity (**d**). Raw data can be presented as scatter plots (**e**, **f**) or histograms (**g**), MFI fold changes of large data sets are best presented as heat maps (p-S6rp is displayed on a different scale, marked in *gray*) (**h**) or overlaid curves (**i**). (**a**–**f)** were plotted using FlowJo v10, (**g** and **h)** using Cytobank and (**i)** using GraphPad Prism. Alexa Fluor 647-conjugated phospho-specific antibodies: Mouse IgG1, κ isotype control, CD3ζ (pY142), SLP76 (pY128), ZAP70 (pY319)/Syk (pY352), Erk1/2 (pT202/pY204), histone H3 (pS10), phospho-tyrosine pY100, S6 ribosomal protein (pS235/236)

- Display the *single-cell* (*SSC*) population in a Pacific Blue/SSC-A density plot and select the different fluorescent barcoding populations based on their Pacific Blue staining intensity (*see* Fig. 2d) (*see* **Note 22**).
- Plot the phospho antibody channel against the FCB channel (*see* Fig. 2e, f) or as histograms (*see* Fig. 2g) to display the phosphorylation events.

4. Calculate the median fluorescent intensity (MFI) fold changes for the different time points by dividing the MFI of the phospho antibody channel of the different FCB populations by the respective MFI of the 0 min (unstimulated, lowest Pacific Blue staining intensity) population. This will set the MFI fold change of the 0 min to 1.
5. MFI fold changes of the phospho antibody channel can then be displayed in various formats like heat maps (*see* Fig. 2h) or line diagrams (*see* Fig. 2i).

3.6 Additional Protocol: Antibody Titration

To obtain optimal results, all antibodies used for flow cytometry, especially if they are phospho specific, should be titrated to determine the correct dilution. Furthermore, many (often un-conjugated) phospho-specific antibodies are available that have been certified by the vendor for the use in other methods than flow cytometry, like western blotting or immune histochemistry. The below described titration protocol tries to address both issues; finding the correct antibody concentrations as well as testing the specificity of the antibodies. Cells are therefore treated ("stimulated") with two phosphatase inhibitors, pervanadate and calyculin A, and a mixture of two signaling activators, PMA and ionomycin. Pervanadate is an inhibitor of protein tyrosine phosphatases and therefore causes the accumulation of phosphorylated tyrosines, whereas calyculin A inhibits serine/threonine phosphatases leading to accumulation of phosphorylated serines and threonines. In contrast to the inhibitors, PMA is activating protein kinase C (PKC). Ionomycin is a Ca^{2+}-ionophore releasing it from the endoplasmic reticulum to the cytoplasm. Thereby, downstream signaling of PKC and Ca^{2+} signaling, which also involves activation of PKC, are activated. The three different samples plus one untreated sample are then fluorescently barcoded, combined, and stained with different concentrations of the antibodies of interest. Antibodies directed against phosphotyrosine epitopes then should only give a signal in the pervanadate, but not in the calyculin A sample, while for phospho-threonine/serine detecting antibodies the opposite should be the case. They should only detect the calyculin A and not the pervanadate-treated cells. Additional information is provided by the PMA/ionomycin. This population should only show increased fluorescence in the phospho-antibody channel, if the signaling molecule is downstream of PKA or induced by Ca^{2+} signaling.

1. For the titration of 15 antibodies, place 4 1.5 mL test tubes, each containing 100 µL of a 5×10^7 cells/mL cell suspension in complete cell culture medium, in a water bath and incubate at 37 °C for 15 min.
2. Add 100 µL fixative to the first tube, mix well by pipetting up and down and incubate for 10 min.
3. Stimulate the second tube by adding each 2 µL PMA and 2 µL ionomycin (100× solutions). Mix well by pipetting up and down and incubate at 37 °C for 10 min.
4. Stimulate the third tube by adding 2 µL calyculin A (10 µM). Mix by pipetting and incubate at 37 °C for 10 min.
5. Stimulate the fourth tube by adding 2 µL pervanadate (100× solution). Mix by pipetting and incubate at 37 °C for 5 min.
6. Add 100 µL of fixative to tubes 2, 3, and 4 after the respective incubation time and incubate at 37 °C for another 10 min.
7. Washing steps, fluorescent barcoding and permeabilization as described in Subheadings 3.1, **step 11**, 3.2 and 3.3 (*see* **Note 23**). Note that all volumes should be doubled since the initial cell suspensions are 100 µL instead of 50 µL.
8. Make serial dilutions of the antibodies in consecutive wells of a 96-well v-bottom plate starting with the concentration recommended by the vendor (*see* **Note 24**). Prepare 100 µL of a 2× solution of the highest antibody concentration in wash solution in a separate tube. Transfer 50 µL into the first well. Then add 50 µL wash solution to the remaining 50 µL in the tube, mix by pipetting and transfer 50 µL to the next well of the plate. Create a series of 4–6 1:1 dilutions continuing this way.
9. Wash, resuspend and stain the cells as described in Subheading 3.4 and run them on the flow cytometer.
10. Analysis of the FCS files and gating strategy as described in Subheading 3.5.
11. Evaluation: The antibody must be specific towards the amino acid and the position in the signaling network. If that is fulfilled, the best separation between background (untreated) and signal defines the optimal antibody dilution (*see* Fig. 3).
12. In addition to a pure visual evaluation, the formula for the *Stain Index*, normally used to calculate the effective brightness of fluorochromes by comparing stained and unstained samples can be used (*see* Fig. 4). Stain Index = D/W (D is the difference between the two MFIs and W in that case is the peak width, equaling 2 × the Standard Deviation, of the unstimulated population).

3.7 Additional Protocol: Adaption to Adherent Cells: Cold Trypsinization

Due to the nature of flow cytometry requiring cells in suspension at the time of analysis, the method described so far is restricted to a relatively limited number of suspension cell lines and to primary samples like PBMCs (peripheral blood mononuclear cells).

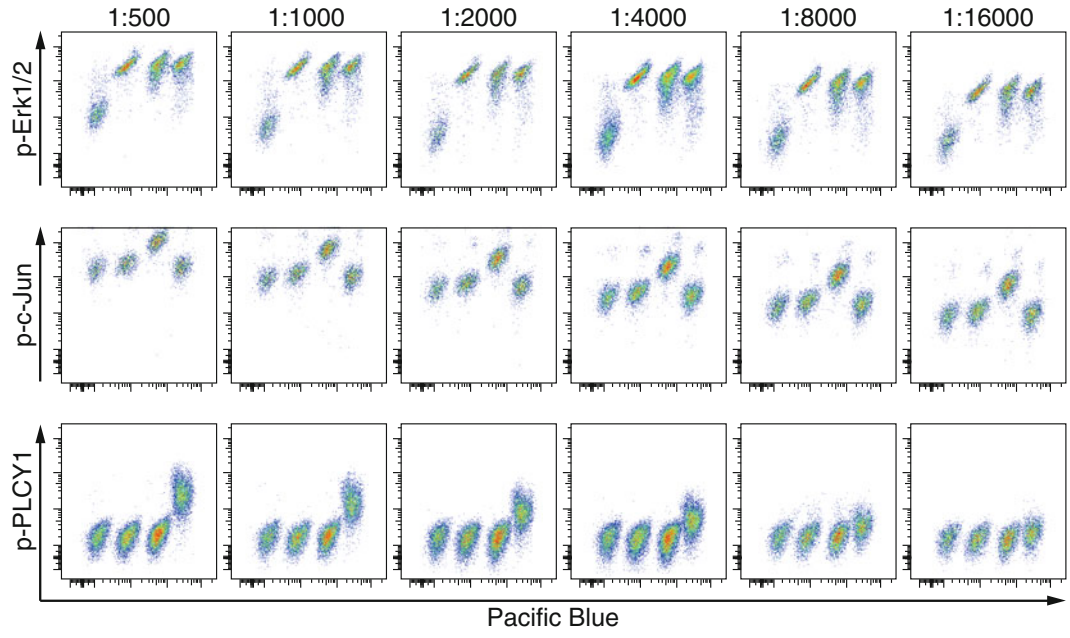


Fig. 3 Antibody titration and specificity test. Staining with a dilution series, final antibody dilutions displayed on top, of three phospho-specific unconjugated antibodies, upper row: Erk1/2 (pT202/pY204), middle row: c-Jun (pT91/pT93) and lower row: PLCγ1 (pY783). FCB populations from left to right: untreated, PMA/ionomycin, calyculin A and pervanadate. As expected, the p-Erk1/2 (threonine, tyrosine, and downstream of PKC) gives a signal in all three, p-c-Jun (threonine, not associated with PKC signaling) only in the calyculin A and p-PLCγ1 (tyrosine, upstream of PKC) only in the pervanadate-treated population. For p-Erk1/2 and p-c-Jun, the background staining is very high but decreases with decreasing antibody concentration. Signal-to-background ratio is largely unaffected by the dilution series. Optimal dilution for both antibodies is 1:16,000. In contrast, p-PLCγ1 displays no background staining, but the signal decreases with the antibody concentration. Optimal dilution is 1:500

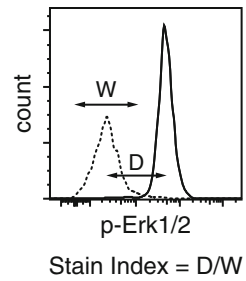


Fig. 4 The Stain Index (SI) can be a good metric for the signal over background or basal phosphorylation. The plot displays histograms of the p-Erk1/2 (1:16,000) signals of untreated (*dashed line*) and PMA/ionomycin-treated cells (*black line*) from Fig. 3. SI = 4.9

However, it is possible to adapt the phospho flow protocol to most adherent cell lines by bringing the cells into suspension prior to flow cytometry. Cells are stimulated in plates at 37 °C, immediately transferred to ice and also trypsinized on ice, which will preserve the phosphorylation status during the trypsin treatment. Afterwards, cells are fixed once they are in solution and can then be treated as described above.

1. Grow cells in a 6-well plate. Each well will represent one time point or condition.
2. Place the plate in a water bath and incubate at 37 °C for 15 min (*see* **Note 8**).
3. Stimulate/treat cells as preferred. In case of time kinetics, the stimulation time course has to be inverted, starting with the stimulation of the cells for the longest time point and ending with the shortest in a way that all stimulation intervals finish at the same time.
4. Transfer the plate to ice (*see* **Note 25**) and quickly remove the stimulation media, starting with shortest time point.
5. Wash the wells with 1 mL of ice cold PBS, to remove proteins from the medium and to improve cooling.
6. Add 700 µL ice-cold 2.5 % trypsin to each well and incubate on ice for up to 30 min (*see* **Note 26**).
7. Quench reactions by adding 600 µL ice-cold complete medium or PBS supplied with 10 % serum and bring cells into suspension by carefully pipetting up and down.
8. Transfer the cell suspensions to a 1.5 mL reaction tube and centrifuge with 300 × *g* at 4 °C for 10 min.
9. Remove supernatants and resuspend cells in 100 µL ice-cold PBS.
10. Add 100 µL fixative to each tube and incubate at 37 °C for 10 min.
11. Continue as described above, starting at step Subheading 3.1, **step 11**.

4 Notes

1. The main ingredient of fix buffer I is paraformaldehyde which is toxic (inhalation and skin contact). Handle with care!
2. The main ingredient of perm buffer III is methanol which is toxic (inhalation and skin contact) and flammable. Handle with care!
3. At http://www.cytobank.org/facselect/ a comparison between different fixation/permeabilization procedures and their compatibility with various antibodies can be found.

4. In combination with stimulation with OKT3 unconjugated primary antibodies should be rabbit derived (poly- or monoclonal, the latter may be preferable if available to target). Mouse antibodies are not advisable in this experimental setup, because the stimulating antibody OKT3 is mouse IgG2a and will therefore be detected by a mouse-reactive secondary antibody. The same will apply for additional cell surface staining.
5. Store the PMA, ionomycin and calyculin A stock solutions in small aliquots at −20 °C. Avoid all too frequent freeze–thaw cycles.
6. The 100× pervanadate solution has to be prepared fresh for each experiment. Vanadate and hydrogen peroxide stock solutions can be stored at 4 °C.
7. Procedures are only described for a single well. However, using eight-channel multi-pipets or robotics will enable for multiple, simultaneous stimulations in the same plate under different conditions, like varying stimulating agents, addition of inhibitors et cetera.
8. Place the plate on a rack in the water bath, ensuring that at least the bottoms of the wells are in good contact with the water. Drilling a small hole in the plate's top in one corner will prevent floating of the plate.
9. If primary T cells are stimulated, co-stimulation via CD28 and/or CD2 is required in addition for a full CD3/TCR activation. For this, cells are preincubated with 1 μg/mL biotinylated OKT3, 5 μg/mL biotinylated anti-CD28, and/or 5 μg/mL biotinylated anti-CD2 at 37 °C for 2 min to allow antibody binding. The time course is then started by cross-linking with 50 μg/mL avidin [6].
10. Optimal fixation conditions are 37 °C for 10 min, but the fixation time can be slightly extended. Therefore, cells of the time points 0 min, 1 min and 5 min can be fixed in the same fixation plate. However, the last, 15-min time point has to be treated separately and can be fixed in the stimulation plate.
11. For longer stimulations, several parameters will have to be adjusted. The cell density has to be lower. A cell culture incubator with 5 % CO_2 atmosphere has to be used instead of a water bath. Cells may change their morphology, which can influence FSC, SSC, auto fluorescence and protein expression levels. It therefore may even become impossible to compare the readouts of unstimulated cells with readouts after too long (days) stimulations.
12. The 10 mg/mL Pacific Blue solution can be stored in small aliquots at −20 °C for up to 1 year. The 100 μg/mL solution can also be reused as multiple freeze–thaw cycles do not destroy the reagent. When thawing, open reagents after they reached

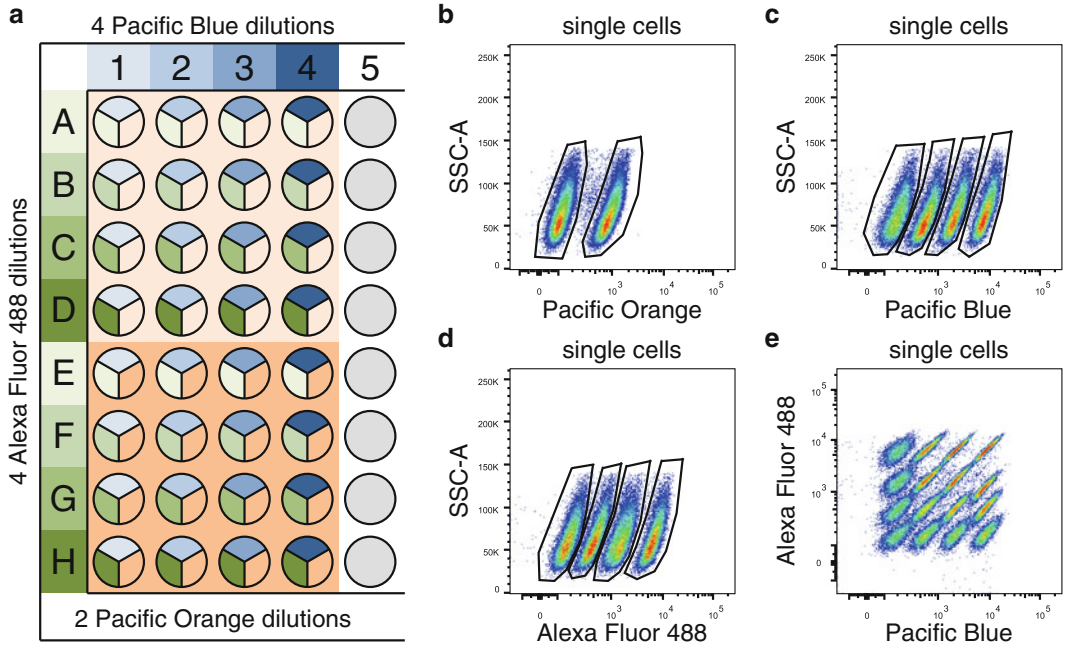


Fig. 5 Multidimensional FCB. (**a**) Layout of a 3D FCB staining matrix in a 96-well plate using four staining intensities of Pacific Blue (*blue*) and Alexa Fluor 488 (*green*), respectively, and two of Pacific Orange (*orange*) to obtain 32 unique combinations. Gating can be done separately for each FCB channel (**b–d**) and gates can be combined in the software to create the right populations for the analysis. FCB channels Alexa Fluor 488 and Pacific Blue plotted against each other (**e**). If populations are not overlapping, they can alternatively be selected directly from this type of scatterplot

room temperature to avoid hydration. In addition to Pacific Blue succinimidyl ester, several other amine-reactive fluorochromes are available, like Pacific Orange succinimidyl ester, excited by the violet laser or Alexa Fluor 488 carboxylic acid, 2,3,5,6-tetrafluorophenyl ester, excited by the blue laser. These reagents can be used instead of Pacific Blue succinimidyl ester, e.g. if the flow cytometer is lacking the violet laser, or in addition to create 2D or 3D staining matrices containing up to 64 samples derived from different conditions. Distinct concentrations of the different dyes thereby have to be combined in a way that assures that every sample is stained with a unique combination of dilutions (*see* Fig. 5a).

13. This dilution series will be suitable for most cell types and also various cell concentrations. It works from comparably small peripheral blood mononuclear cells (PBMSs) to much larger adherent cell lines.

14. The reagents are highly reactive. Direct mixing therefore is crucial for a homogenous staining of the entire cell suspension. Always add the cells to the FCB reagents. Adding the reagents to the cells will not result in distinct populations.

15. Proper resuspension before adding the permeabilization solution is crucial to avoid clumping and formation of aggregates through the methanol.
16. Samples can be stored in the permeabilization solution at −80 °C for several months.
17. The density of the cell suspension should be approximately in the range of the cell density used for the antibody titration. In our laboratory usually between 0.5 and 2×10^6 cells/50 μL. For significantly higher densities, the total staining volume should be adjusted.
18. Having all phospho-specific antibodies conjugated to the same bright fluorochrome, here Alexa Fluor 647, leaves the other channels free for FCB and surface marker staining, and makes later on analysis easier. In order to reduce compensation issues in the phospho-antibody channel, it is advisable to leave the other channels on the same laser unused, if possible.
19. The staining panel should always contain at least one isotype control. Antibodies that detect the unphosphorylated forms of the proteins are not necessary. Total protein levels should normally not change during short stimulations. Cell surface marker staining can be done in the same mix. Note that some epitopes will not be recognized after methanol treatment. These stainings can be performed before the permeabilization. A list of cell surface markers and buffer compatibility can be found at: http://www.bdbiosciences.com/documents/antibodies_human_cellsurface_marker.pdf
20. Spillovers between Alexa Fluor 647 and Pacific Blue are in both directions very small (between 0.00 % and 0.10 %) because they are excited by different lasers.
21. The scattered light of particles passing through the lasers is recorded by photomultipliers (PMTs) as voltage pulses having a peak height, width, and area. The width correlates with the particle size and can therefore be used during the analysis to discriminate doublets or larger aggregates based on the disproportion between height and width compared to single cells, especially in the side scatter (SSC).
22. In principle, the barcoding populations can be gated based on any scatter plot having the barcoding dye signal on one axis. However, displaying the barcoding signals against the SSC-A usually provides the best separation. Gating based on histograms is not advisable. In the case of multidimensional FCB, every barcoding channel can be gated separately against the SSC-A (*see* Fig. 5b–d). The final populations can then be created by combining the different gates using *Boolean gating* (&) in FlowJo or in Cytobank by adding populations in the population manager and assigning the correct gate combina-

tions to the new populations. For 2D barcoding, it might also be possible to directly gate the populations based on a scatter plot of the two FCB channels against each other (*see* Fig. 5e), if there is no major overlap between populations. If cell surface markers are used in addition to FCB and more than one population is examined, the use of Cytobank is recommended. There, the populations from surface markers can be used as a separate layout dimension in the illustration/analysis, which makes the analysis more convenient.

23. The total number of 20 million cells is sufficient for the titration of ~15 antibodies. If fewer antibodies are evaluated, the cell mixture can be aliquoted during the permeabilization step and stored at −80 °C for later titrations.
24. In case of unconjugated antibodies a good starting point (highest concentration) for the titration is a final dilution of 1:100.
25. Make sure that the plate is leveled and there is good contact between the ice and the bottom of the plate. A plate filled with warm water can be used to adjust the ice surface beforehand.
26. The time span for the cold trypsinization should be evaluated in advance, using, e.g., a 24- or 48-well plate and trypsinization intervals from 5 to 30 min. Not all cell lines are suitable for cold trypsinization, because some adhere too well or grow in microspheres.

Acknowledgements

Our research is funded by grants from the Research Council of Norway, Norwegian Cancer Society, the European Commission (7th Framework Programme), South Eastern Norway Regional Health Authority, the K.G. Jebsen Foundation, and Novo Nordic Foundation which has permitted the method development as described here. Johannes Landskron is a Senior Scientist funded by the Jebsen Centre for Cancer Immunotherapy.

References

1. Krutzik PO, Nolan GP (2003) Intracellular phospho-protein staining techniques for flow cytometry: monitoring single cell signaling events. Cytometry A 55(2):61–70. doi:10.1002/cyto.a.10072
2. Oberprieler NG, Tasken K (2011) Analysing phosphorylation-based signalling networks by phospho flow cytometry. Cell Signal 23(1):14–18. doi:10.1016/j.cellsig.2010.07.009
3. Storvold GL, Landskron J, Strozynski M, Arntzen MO, Koehler CJ, Kalland ME, Tasken K, Thiede B (2013) Quantitative profiling of tyrosine phosphorylation revealed changes in the activity of the T cell receptor signaling pathway upon cisplatin-induced apoptosis. J Proteomics 91:344–357. doi:10.1016/j.jprot.2013.07.019
4. Krutzik PO, Clutter MR, Trejo A, Nolan GP (2011) Fluorescent cell barcoding for multiplex

flow cytometry. Current protocols in cytometry/editorial board, J Paul Robinson, managing editor [et al] Chapter 6:Unit 6 31. doi:10.1002/0471142956.cy0631s55

5. Krutzik PO, Nolan GP (2006) Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. Nat Methods 3(5):361–368. doi:10.1038/nmeth872

6. Kalland ME, Oberprieler NG, Vang T, Tasken K, Torgersen KM (2011) T cell-signaling network analysis reveals distinct differences between CD28 and CD2 costimulation responses in various subsets and in the MAPK pathway between resting and activated regulatory T cells. J Immunol 187(10):5233–5245. doi:10.4049/jimmunol.1101804

Part VI

Bioinformatics Resources for Phosphoproteomics Analyses

Chapter 20

Resources for Assignment of Phosphorylation Sites on Peptides and Proteins

Vaishnavi Ravikumar, Boris Macek, and Ivan Mijakovic

Abstract

Reversible protein phosphorylation is a key regulatory posttranslational modification that plays a significant role in major cellular signaling processes. Phosphorylation events can be systematically identified, quantified, and localized on protein sequence using publicly available bioinformatic tools. Here we present the software tools commonly used by the phosphoproteomics community, discuss their underlying principles of operation, and provide a protocol for large-scale phosphoproteome data analysis using the MaxQuant software suite.

Key words Bioinformatics, Phosphorylation site identification, MaxQuant

1 Introduction

Cellular functions of proteins are often tightly regulated by their respective posttranslational modifications (PTM). Identification and quantification of modified proteins is important for the understanding of signal transduction networks, whereas the knowledge of the precise site of modification on the protein sequence is critical for a mechanistic understanding of the influence of the PTM on individual protein function [1, 2]. Experimental identification of phosphorylation sites is relatively labor intensive; hence significant efforts have been invested in constructing *in silico* phosphorylation site predictors. Different algorithms have been implemented to construct predictors based on experimental datasets [3, 4] and protein structural features [5]. Since phosphorylation sites are usually not well conserved among species, predictors have also been constructed based on clade-specific data [6–8]. While this field is definitely showing progress, the performance of *in silico* predictors is still not sufficient for practical use [9], and experimental identification of phosphorylation sites remains at present the only viable alternative.

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_20,

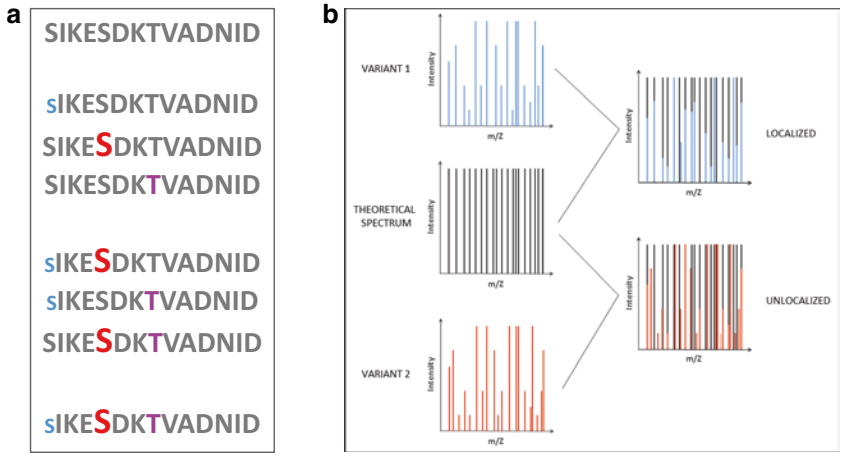
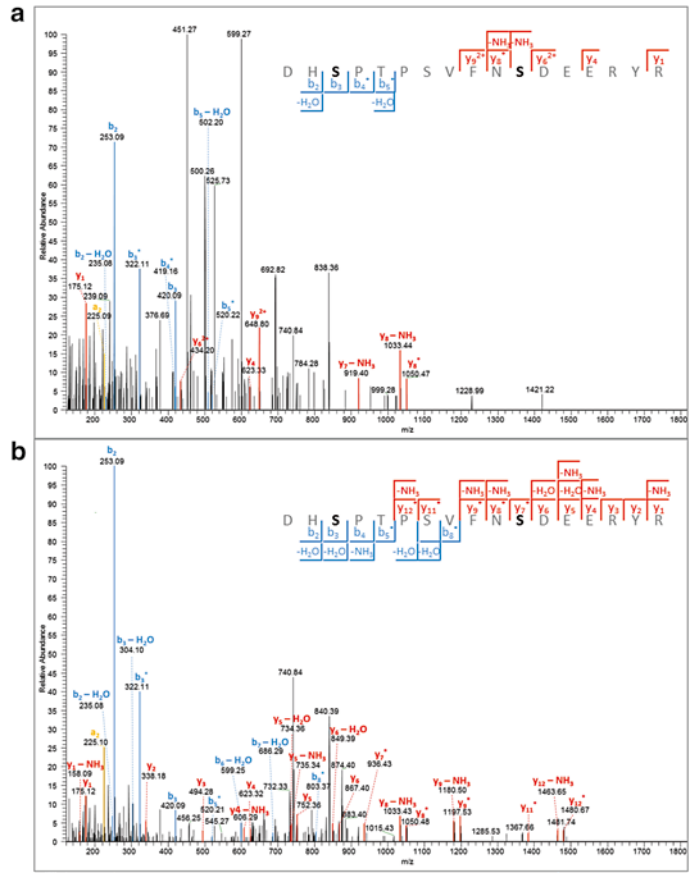


Fig. 1 (**a**) A peptide can have a number of variants depending on the presence of number of residues that can be modified, calculated as 2^n. In the above shown example, eight variants of the peptide are possible, including the unmodified and phosphorylated forms (singly, doubly, or triply). The probability of localization of the phospho-modification to a particular amino acid is represented by the letter size as an example. (**b**) An example of a peptide spectrum match and assignment. Variant 1 matches the theoretical spectrum better than Variant 2

During the past two decades, mass spectrometry has emerged as a method of choice to study PTMs and signal transduction networks. Traditionally, phosphoproteomics relied on 2D gels to identify and quantify phosphoproteins [10]. Those studies, although capable of visualizing and identifying hundreds of phosphoproteins on gels, could rarely retrieve information on the exact localization of phosphorylation events. However, current large-scale, quantitative phosphoproteomics studies based on biochemical enrichment of modified peptides and "shotgun" MS protocols routinely detect >10,000 phosphorylated peptides in a typical eukaryotic cell line in a single experiment [11, 12]. Since the peptide sequence can be easily inferred from the MS/MS fragmentation pattern, these new protocols enable straightforward localization of many modification sites. This review focuses on the present day bioinformatics tools routinely used to confidently localize phosphorylation sites on a peptide or protein.

Protein/peptide identification relies on search engines that match peptide fragmentation patterns to a database of known protein/peptide sequences [13]. While determination of the peptide sequence is relatively straightforward, detection and localization of modification sites are more complicated. One reason for this is the fact that presence of PTMs on peptides increases the number of theoretical spectra that need to be matched to a mass spectrum (search space) (*see* Fig. 1a). For example, if a particular peptide is doubly phosphorylated and possesses three modifiable amino acids (Ser, Thr, or Tyr), it will have eight possible PTM variants that will



Information Panel

Scan Number	12622
Gene Name	FIP1L1
Mass	2194.8358
m/z	549.71623
Position in Protein	490,506
Localization prob.	0.956,0.943
PEP	4.9167E-05
Score	52.483
Charge	4
Retention Time	60.002

Information Panel

Scan Number	12572
Gene Name	FIP1L1
Mass	2194.8358
m/z	732.61921
Position in Protein	490,506
Localization prob.	0.999,1
PEP	2.1037E-63
Score	158.02
Charge	3
Retention Time	59.778

Fig. 2 Good sequence coverage and low background noise is important for accurate peak annotation. Shown here is an example of (**a**) bad and (**b**) good MSMS spectrum of a phosphopeptide along with their respective information panels

have to be matched to the mass spectrum in order to pinpoint the exact location of the modifications. Increased search space as a consequence of PTM consideration during database search may also largely escalate false-positive identifications. Furthermore, localization of a modification on a peptide requires comprehensive MS fragmentation patterns, but the presence of the phosphorylation modification often interferes with fragmentation resulting in lower coverage and significance scores (*see* Fig. 2). For example, loss of the modification from the peptide bond is often the main fragmentation event, especially for modifications occurring on Ser/Thr residues such as O-glycosylation and phosphorylation (so-called neutral loss). In these cases, CID often results in a prominent unmodified precursor ion peak that is devoid of information on peptide sequence and additional fragmentation needs to be performed on that peak for sequence and modifications site assignment [14]. Since Ser/Thr phosphorylation is "lost" in the form of

phosphoric acid, the modified serine is converted into dehydroalanine and the resulting mass shift (−18 Da) can be used to localize the modification site. Conversely, the localization of O-glycans is more challenging, as they are lost without changing the backbone sequence and therefore the information on their exact location is missing. This problem can be addressed by specific (prolonged) CID acquisition regimes [15] or the use of other fragmentation techniques, such as electron capture dissociation (ECD) [16] or electron transfer dissociation (ETD) [17] that preserve labile modifications on the peptide backbone. Although not of fundamental importance, high measurement mass accuracy (MMA) is also important in analysis of certain modifications. Besides its generally positive influence on selectivity in case of large database search space, high MMA is needed to resolve modifications that are very similar in mass, such as tyrosine phosphorylation (79.966 Da) and sulfation (79.956 Da); acetylation (42.011 Da) and tri-methylation (42.047 Da); and formylation (27.990 Da) and dimethylation (28.031 Da). This is also important in the context of cross talk between different modifications, as an increasing number of studies report interplay of different modifications situated in close proximity on proteins [18, 19].

For localization scoring purposes search engines such as Mascot or Sequest have integrated PTM-scoring algorithms, such as Delta-score or *A*-score, respectively. Software suites such as MaxQuant have integrated scoring algorithms (PTM Score), in addition to other modules for MS data processing, search engine and statistical post-processing of MS data. Additionally stand-alone programs such as PhosphoRS can also be used which can apply post-raw data processing with any software. MS/MS data interpretation can be done with various additional MS vendor tools such as Proteome Discoverer (Thermo Scientific), ProteinLynx (Waters), MassHunter/Spectrum Mill (Agilent), and others.

One software tool that uses high MMA to find and map PTMs is ModifiComb [20]. A base peptide which is the unmodified counterpart of a particular peptide and a dependent peptide which is the modified or mutated counterpart of the same peptide is assumed to be present in the same sample mixture and LC-MS run forming a so-called peptide family. These are expected to elute within a limited time window of an LC-MS run, either before or after the other. Thus the algorithm encoded within ModifiComb looks for peptides with strong sequence similarities and calculates the ΔM between their molecular masses and ΔRT between their retention times. It builds a histogram based on the ΔM and ΔRT values and identifies pair matches for a minimum of four dependent peptide fragments coinciding with the base peptide fragments, within a 20 mDa window. Generally the base peptide is identified by de novo sequencing or database search whereas the dependent peptide should not be identified unless the variable

modification is specified during the search. Likewise, dependent peptide analysis incorporated within the Max Quant search algorithms can also be used to identify modified peptides using the information obtained from their identified unmodified analogues [21]. This is done by calculating a probability score for every unidentified peptide based on its match to every identified peptide in order to find the closest or best match to its base peptide. The Δm information of the best scored peptide is then used to assign the modification and position. Error tolerant search [22], with Mascot, is yet another tool for finding unsuspected modifications on a peptide. It consists of a standard first search that uses predefined search parameters. Identified proteins are then selected for a second error tolerant search with no enzyme specificity and a complete and comprehensive list of chemicals and posttranslational modifications. The identified modified peptides are then scored to mark the best matches.

Despite the acknowledged fact of the importance of PTM localization, there is a lack of one complete solution to all the above-described issues. However there are some fundamental strategies widely used in the field to help to overcome some of these obstacles.

1.1 Phosphorylation Site Localization Scoring Strategies

Previously, site localization was assessed by manual verification of fragmentation spectra, making site assessment a labor-intensive and time-consuming procedure. With the development of various software tools, phosphorylation site localization has become automated and more accurate. PTMs on peptides are detected as a mass increment or deficit relative to the respective unmodified form of the peptide. Presence of a PTM on peptide fragments causes the *m/z* of the peptide to shift and these modified and unmodified fragments are of special importance in localization; absence of these fragments results in ambiguity. Algorithms use this information to map the peptide to the best possible spectrum and generate a probability-based localization scoring for phosphorylation assignment (*see* Fig. 1b). Contrary to assessment of *f*alse *d*iscovery *r*ate (FDR), for site localization there is no similar equivalent measure or *f*alse *l*ocalization *r*ate (FLR). An incorrectly localized site would not be considered as a decoy match but have a close match to the correct spectrum [23]. Thus probability based scoring algorithms are used for this purpose. The proteomics field commonly makes use of localization scoring methods primarily developed by two groups—the *A*-score by Beausoleil *et al.* [24] and the PTM score by Olsen *et al.* [25, 26] or modified variants of these.

1.1.1 PTM Score

The PTM score was developed for localizing phosphorylation site identifications in low mass accuracy ion trap-collision-induced dissociation data. It gives the probability of a match that is directly calculated from the MS^3 spectrum to the respective fragment ion sequence of a particular peptide. In the original publication, the

top four most intense signals were chosen by the algorithm for a given fragment for every 100 Da window. This value, chosen based on experimental observations, was also corroborated by a group in Israel [27] who noticed that peaks from two M/MS spectra correspond to each other if their masses do not differ more than 0.4 atomic mass units (amu). They used the most intense peak for the match as implemented in Pep-Miner. Thus the probability of the calculated masses to match the experimentally recorded masses is 4/100 or 0.04 and this is independent of the mass range under consideration. In more recent publications, the peak number was increased to 10 [28]. The reduced mass list creates a satisfactory balance between true positives and background noise. Next, the algorithm calculates the b- and y-ions and their respective internal cleaved product ion masses for potential b- and y-ion precursors and determines the number of matches, denoted as k. The probability P of obtaining random matches between computed and measured peaks is calculated using

$$P = \binom{n}{k} . p^k . (1-p)^{(n-k)}$$

where P is the probability of obtaining k successful matches in n number of fragments for a particular mass range. This algorithm thus calculates the probability of each potential modified site and then sums up all probabilities to 100 %. This normalized scale makes the comparison between different potential sites on the same peptide convenient.

Recorded peptide fragment spectra are scored automatically with the algorithm encoded within the MaxQuant [21] software suite, which is an open-source software available for download at http://www.maxquant.org/index.htm.

1.1.2 A-Score

This algorithm was also developed to assess low accuracy ion trap data. Similar to the PTM score strategy, this algorithm divides the m/z range into 100 bins and from each bin the six most intense peaks are chosen to generate a peak list for scoring purposes. The probability score is calculated based on the presence and intensity of site determining ions that are exclusive to a specific site, in the MS/MS. The difference of the A-score to the PTM score lies in the fact that the A-score does not report the probability score of all possible sites but only of the best one. This is calculated as the score difference between the best site and the next best and finally reported as 10 $\log_{10}$ p-score. Phosphorylation site localization by A-score is determined by identifying the possible sites of phosphorylation followed by probability calculation of the phosphorylation site localization based on the likelihood of identifying site determining ions rather than by chance. The probability of predicted b- and y-ions matching an identified spectrum is calculated by the equation

$$P(X) = \sum_{k=n}^{N} \binom{N}{k} p^k (1-p)^{(N-k)}$$

where P represents the probability of the fragment ions randomly matching to the MS^2 spectrum in N the total number trials with (n) number of successful events, p being the probability of success and represented as $-10 \log (P)$.

This algorithm is integrated with the Sequest [29] software and the matched peptide sequences are given a probability based ion matching score (also called peptide score). The threshold score used by the developers to validate their scoring algorithm [24] was $p < 0.01$. Apart from providing information regarding the precise site of localization, *A*-score also provides information about the unlocalized modification sites.

1.1.3 Mascot Delta Score

The Mascot Delta score encompasses a reevaluated scoring strategy described by Savitski *et al.* in 2011 [30]. The Mascot Delta score or MD score concept has been tested and is applicable to many fragmentation techniques such as CID, MSA, ETD, and HCD. It was developed for use with the Mascot [31] database search engine. MD scores are calculated separately using mascot. dat as an input file. The MD score algorithm was formulated to objectively assess phosphorylation site assignments made by Mascot. Contrary to the PTM score or *A*-score where the localization is a result of post database search re-analysis of peptide spectral matches, MD score makes use of native Mascot scores to calculate uncertainty in the reported localization scores. The straight MD score (as opposed to normalized MD score previously tested by Beausoleil *et al.* [24]) is calculated as a ratio difference between the top two Mascot ion scores of alternative phosphorylation sites in a given peptide sequence. Since only the best ion score is considered, MD score calculation utilizes the peptide score of either the one containing the phosphorylated amino acid or the one containing the neutral loss of the phosphoric acid. As in case of the above two discussed strategies, MD scoring algorithms also divide the *m/z* range into fragment windows, the difference from the above two being that it is divided into 110 bins and "n" most intense peaks are then chosen. Additionally, the MD score does not show a strong bias towards any of the different amino acids (namely S/T/Y), although it is less sensitive for p-S/T peptides. Furthermore, Savitski *et al.* also report different MD score thresholds for different fragmentation techniques making assignment of phosphorylation sites possible irrespective of the nature of instrument or fragmentation technique used. Finally, the MD scoring approach can be used to score large as well small datasets as the scoring is independent of the size of the database searched against.

1.1.4 Additional Tools

In addition to the above described commonly used tools for the precise localization of sites identified from phosphoproteomics datasets, numerous other programs have been developed for the purpose of overcoming the shortcomings of the existing ones. The key features of some of these software tools are briefly described below.

ModLS is a variant of the PTM score but is designed for localizing many PTM types and considers entries from UniMod. In addition it can be used to analyze a dataset from multiple search algorithms and can combine their results. It has been described as "a user-friendly localization tool for arbitrary PTMs" [32]. In addition, ModLS also has provisions for identifying and minimizing mis-assigned phosphorylation sites.

PhosphoRS [33] localization software can also be used in conjunction with all common fragmentation techniques and with high and low-mass-accuracy data. It is a stand-alone software but can also be installed within the Thermo Scientific Proteome Discoverer software. It is downloadable from http://cores.imp.ac.at/protein-chemistry/download/. The combination of a novel peak extraction procedure (flexibility in number of peaks chosen per 100 *m/z* window) and the optimization of scoring criteria for various fragmentation methods (accounting for singly and doubly charged fragment ions and neglecting neutral loss peaks) makes it a unique localization tool. Further, the authors also show the practicality of this software by implementing it to localize sites identified from HeLa cells after TiO_2 enrichment.

SLIP (or Site Localization In Peptide) [34] scores all peptides identified by the Batch-Tag search engine in Protein Prospector (http://prospector.ucsf.edu/prospector/mshome.htm). The algorithm picks the 40 most intense peaks in each *m/z* range. An important point to note is that this algorithm can be used for high-mass-accuracy data for site localization. SLIP scoring lists all potential phosphorylated sites within the set score threshold and considers other ion types and is not restricted to only b/y ions.

SLoMo (or Site Localization of Modifications) [35] is an adaptation of the widely used *A*-score and is designed specifically to also assess electron transfer dissociation data and can only be implemented with the Sequest and OMSSA [36] proteomics analysis software suite searches. SLoMo recognizes the generic pepXML input format and can be used for any modification found it the UniMod database. MS^2 data are used to calculate specific site probabilities and approximated by a Poisson distribution.

2 Protocol: Analysis of Phosphorylation Datasets Using MaxQuant

Described below is a step-by-step procedure for processing and analyzing phosphoproteomics data using the MaxQuant software suite.

1. Load Xcalibur.raw files in MaxQuant (use the latest version - here we describe v. 1.5.0.0) under the "Raw files" tab using the Load/Load folder function.

Note 1: All files to be analyzed should be copied into a single folder on your local computer before analyzing them with MaxQuant. Any version of MaxQuant can be used to specifically process Xcalibur.raw files generated by Thermo Fischer Instruments. In addition, the newest 1.5.0.0 version can also support raw data from AB Sciex TripleTOF 5600 and Bruker Impact HD.

2. The "write template" function will create a combined folder in the same folder that contains all copied raw files to be analyzed. The experimentalDesignTemplate.txt file in the combined folder can be defined and modified accordingly.
3. Under the "Group-specific Parameters"—General tab, choose "Phospho (STY)" as a variable modification.
4. Specify the endoprotease and label used under the same tab, maintaining all other default parameter settings.
5. Load the protein sequence database of choice (downloaded from UniProt for example) under the "Global parameters"—General tab by selecting the "Add file" function.

Note 2: Databases can be obtained from different sources like UniProt, RefSeq/NCBI, or other specialized platforms for specific organisms such as TAIR for Arabidopsis. These databases obtained from different sources differ mainly with respect to the protein identifiers. Essentially any of the above mentioned databases can be used.

6. The Advanced tab under the "Global parameters" has provision for selection of the combined folder destination.
7. Choose the appropriate number of threads and select Start. The processing time is generally dependent on the number of raw files and search space.

Note 3: Number of threads refers to the number of physical cores of the computer and it should be the minimum value of the number of cores and number of raw files combined. For example two threads should be used if only two raw files are being processed on a six-core computer; or six threads should be used if 10 raw files are being processed on a six-core computer.

Note 4: A more detailed protocol on the requirements and usage of MaxQuant can be obtained from [37] or from http://maxquant.org/requirements.htm.

8. On completion, all processed data can be found in the same folder as the raw files, in a combined\combined\txt folder as .txt files which can be conveniently read and edited with MS Office-Excel.

9. Based on the modification defined, a generically labeled .txt file is created after every search, for example; Phospho (STY)Sites.txt or Acetyl(K).txt. Specifically, in case of the phospho-modification, based on the modified residue mentioned during database search, the file will be labeled Phospho (STY)Sites.txt for modification of residues Ser/Thr and Tyr or just Phospho (S)Sites.txt for modification on only Ser, for example.
10. The Phospho (STY)Sites.txt contains all information regarding all identified and/or quantified phosphorylation sites.

    Note 5: During the database search, we also include 248 common laboratory contaminant entries which are then additionally filtered for in the Phospho (STY)Sites.txt.
11. Filter all reverse hits in the Phospho (STY)Sites.txt.

    Note 6: Reverse hits refers to protein hit matches to the reverse sequences from the decoy database.
12. Here we describe some of the most important columns from the Phospho (STY)Sites.txt file. Figure 3 is provided as an example of the MaxQuant output.
    - *Localization prob*—the confidence of the site assignment to specific amino acids can be estimated with the help of this column. It indicates the likelihood of occurrence of the phosphorylation event on a specific amino acid. All sites listed in this table are localized with either a good or bad probability value.
    - *Note 7*: Generally, we consider a localization probability of $p \geq 0.75$ as acceptable in our laboratory.
    - *Proteins*—states the UniProt identifier as well as the gene names of every peptide entry.
    - *Fasta headers*—gives the protein description of every entry.
    - *Position within proteins*—states the position of the phosphorylation site within the protein sequence.
    - *Amino acid*—states the residue that has been identified as modified.
    - *PEP*—or Posterior Error Probability is the score of the corresponding phosphorylated peptide identification. It helps to gauge the statistical significance of the identified phosphorylated peptide.
    - *Score*—is the peptide score for the best identified phosphorylated peptide assigned by the Andromeda search engine.
    - *Ratio H/L* (or *M/L*) *normalized*—gives the required quantitative information. It lists all normalized ratios between two labeled partners where the median of the total ratio population is shifted to 1.

a

Some of the important columns from the original Phospho(STY)Sites table are listed and described here	
Column Header	**Description**
Proteins	UniProt identifier and gene name of the respective protein
Positions within proteins	Position of the identified phosphorylation event within the peptide sequence
Fasta headers	Description of FASTA headers for UniProtKB
Local prob	Localization Probability indicating the likelihood of occurrence of the phosphorylation event on a specific amino acid
PEP	Posterior Error Probability score of the corresponding phosphorylated peptide identification
Score	Peptide score assigned by the Andromeda search engine
Amino acid	The amino acid that has been identified as modified
Ratio H/L Norm	Normalized ratio between two labeled partners (heavy to light) where the median of the total population is shifted to 1.
Ratio H/L Norm__1	Normalized ratio, of peptide variant 1, between two labeled partners (heavy to light) where the median of the total population is shifted to 1.
Ratio H/L Norm__2	Normalized ratio, of peptide variant 2, between two labeled partners (heavy to light) where the median of the total population is shifted to 1.
Ratio H/L Norm__3	Normalized ratio, of peptide variant 3, between two labeled partners (heavy to light) where the median of the total population is shifted to 1.
Occupancy L	Stoichiometry of the modification in the light labeled species, in the case of a labeled experiment
Occupancy H	Stoichiometry of the modification in the heavy labeled species in the case of a labeled experiment
Intensity	Total intensity of all isotopic patterns in the label cluster
Rev	Short for Reverse.'+'denotes that this particular entry had a match to the decoy database
Cont	Short for Contaminant. '+' denotes that this particular entry was identified as a commonly occurring laboratory contaminant

b

Proteins	Position within protein	Fasta Headers	Local. Prob	PEP	Score	Amino acid	Ratio H/L Norm	Ratio H/L Norm _1	Ratio H/L Norm _2	Ratio H/L Norm _3	Occupa -ncy L	Occupa - ncy H	Intensity	Rev	Cont
CON__P01045-1	315	>P01045-1 SWISS-PROT:P01045-1 (Bos taurus) Isoform HMW of Kininogen-2 precursor	0.5	0.0161272	61.48	S	0.39851	0.39851	NaN	NaN	NaN	NaN	8621500		+
sp|O31996|YOKK_BACSU	2	>sp|O31996|YOKK_BACSU SPBc2 prophage-derived uncharacterized protein YokK OS=Bacillus subtilis (strain 168) GN=yokK PE=4 SV=1	0.957395	0.00584903	118.71	S	NaN	NaN	NaN	NaN	NaN	NaN	12364000		
sp|O34660|ALDH4_BACSU	254	>sp|O34660|ALDH4_BACSU Putative aldehyde dehydrogenase DhaS OS=Bacillus subtilis (strain 168) GN=dhaS PE=3 SV=1	0.997605	0.00899239	81.95	S	NaN	NaN	NaN	NaN	NaN	NaN	7902100		
sp|P08877|PTHP_BACSU	46	>sp|P08877|PTHP_BACSU Phosphocarrier protein HPr OS=Bacillus subtilis (strain 168) GN=ptsH PE=1 SV=3	1	1.27E-07	179.66	S	0.018784	0.018784	NaN	NaN	0	0	1996900000		
sp|Q99027|COMP_BACSU	437	>sp|Q99027|COMP_BACSU Sensor histidine kinase ComP OS=Bacillus subtilis (strain 168) GN=comP PE=2 SV=3	0.999996	0.039795	54.09	S	9.0416	9.0416	NaN	NaN	NaN	NaN	89116000		
sp|P71062|EPSL_BACSU	59	>sp|P71062|EPSL_BACSU Uncharacterized sugar transferase EpsLOS=Bacillus subtilis (strain 168) GN=epsL PE=2 SV=1	0.999953	0.010908	70.908	T	0.022772	NaN	0.022772	NaN	NaN	NaN	290160000		

Fig. 3 (**a**) Detailed description of column headers present in the MaxQuant Phospho (STY) Sites table output. (**b**) Example of a MaxQuant Phospho (STY) Sites table output

- *Ratio H/L_1 (2 or 3)*—a particular peptide can be detected as multiple variants based on the number of possible phospho-modification sites on the peptide. The *Ratio H/L_1 (2 or 3)* columns report the ratios between labels for each detected modified variant, respectively.
- *Intensity*—refers to the total intensity value of all isotopic clusters related to the identified peptide. It is the summed up Extracted Ion Current of all isotopes in the labeled cluster.
- *Occupancy*—gives stoichiometric information about the phosphorylation modification and can be defined as the ratio of the phosphorylated and non-phosphorylated version of the same peptide with respect to an identical site.
- *Reverse*—rows marked with a "+" indicate that the respective peptide could be matched to a reverse protein sequence

derived from the decoy database. These must be filtered before data analysis.

- *Contaminant*—rows marked with a "+" indicate that the respective peptide was found to be a part of a common laboratory contaminant. These must be filtered before data analysis.

3 Conclusion

Protein phosphorylation is one of the most prominent and ubiquitous modifications that govern all aspects of a cell. Thus it is crucial to identify and further understand the exact regulatory role of this modification in signaling processes. Advances in technology and enrichment techniques have now made identification of phosphorylation events occurring in a cell a routine. However it is equally crucial to recognize the precise site of the modification on the peptide. This demands the improvement of MS acquisition methods and computational methods for PTM prediction and localization. There has been an impressive development in the advancement of scoring strategies through the years, but there are still certain pitfalls that exist in current approaches. There is a need for more sensitive methods that can accurately localize modification sites and measure FLR without matching of thousands of theoretical fragment masses to MS spectra, e.g., by using already existing or specifically produced spectral libraries. There is also a need for a benchmark dataset that can be used consistently for developing new or improving current localization algorithms. Filling these gaps would help the proteomics community to take one step further towards confidently publishing reliable large-scale PTM datasets in complex systems.

Acknowledgments

This work was supported by grants from the Chalmers University of Technology (to IM), the Juniorprofessoren-Programm of the Landesstiftung BW, the SFB766 of the Deutsche Forshungsgemeinschaft, and PRIME-XS consortium (to BM).

References

1. Roux PP, Thibault P (2013) The coming of age of phosphoproteomics—from large data sets to inference of protein functions. Mol Cell Proteomics 12(12):3453–64
2. Mijakovic I, Macek B (2012) Impact of phosphoproteomics on studies of bacterial physiology. FEMS Microbiol Rev 36(4): 877–92

3. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294(5):1351–62
4. Zhao X et al (2012) Prediction of protein phosphorylation sites by using the composition of k-spaced amino acid pairs. PLoS One 7(10), e46302
5. Iakoucheva LM et al (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32(3):1037–49
6. Ingrell CR et al (2007) NetPhosYeast: prediction of protein phosphorylation sites in yeast. Bioinformatics 23(7):895–7
7. Miller ML et al (2009) NetPhosBac - a predictor for Ser/Thr phosphorylation sites in bacterial proteins. Proteomics 9(1):116–25
8. Que S et al (2012) PhosphoRice: a meta-predictor of rice-specific phosphorylation sites. Plant Methods 8:5
9. Que S et al (2010) Evaluation of protein phosphorylation site predictors. Protein Pept Lett 17(1):64–9
10. Eymann C et al (2007) Dynamics of protein phosphorylation on Ser/Thr/Tyr in *Bacillus subtilis*. Proteomics 7(19):3509–26
11. Huttlin EL et al (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143(7):1174–89
12. Olsen JV et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3(104):ra3
13. Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5(9):699–711
14. Martin DB et al (2005) Investigation of neutral loss during collision-induced dissociation of peptide ions. Anal Chem 77(15):4870–82
15. Gonzalez de Peredo A et al (2002) C-mannosylation and o-fucosylation of thrombospondin type 1 repeats. Mol Cell Proteomics 1(1):11–8
16. Zubarev RA et al (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem 72(3):563–73
17. Syka JE et al (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101(26):9528–33
18. Beltrao P et al (2013) Evolution and functional cross-talk of protein post-translational modifications. Mol Syst Biol 9:714
19. Soufi B et al (2012) Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. Curr Opin Microbiol 15(3): 357–63
20. Savitski MM, Nielsen ML, Zubarev RA (2006) ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. Mol Cell Proteomics 5(5):935–48
21. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–72
22. Creasy DM, Cottrell JS (2002) Error tolerant searching of uninterpreted tandem mass spectrometry data. Proteomics 2(10):1426–34
23. Chalkley RJ, Clauser KR (2012) Modification site localization scoring: strategies and performance. Mol Cell Proteomics 11(5):3–14
24. Beausoleil SA et al (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 24(10):1285–92
25. Olsen JV, Mann M (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc Natl Acad Sci U S A 101(37):13417–22
26. Olsen JV et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635–48
27. Beer I et al (2004) Improving large-scale proteomics by clustering of mass spectrometry data. Proteomics 4(4):950–60
28. Marx H et al (2013) A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. Nat Biotechnol 31(6):557–64
29. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5(11):976–89
30. Savitski MM et al (2011) Confident phosphorylation site localization using the Mascot Delta Score. Mol Cell Proteomics 10(2):M110 003830
31. Perkins DN et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551–67
32. Trudgian DC, Singleton R, Cockman ME, Ratcliffe PJ, Kessler BM (2012) ModLS: post-translational modification localization scoring with automatic specificity expansion. J Proteomics Bioinform 5(12):283–289
33. Taus T et al (2011) Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 10(12):5354–62

34. Baker PR, Trinidad JC, Chalkley RJ (2011) Modification site localization scoring integrated into a search engine. Mol Cell Proteomics 10(7):M111 008078
35. Bailey CM et al (2009) SLoMo: automated site localization of modifications from ETD/ECD mass spectra. J Proteome Res 8(4):1965–71
36. Wenger CD et al (2011) COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. Proteomics 11(6):1064–74
37. Cox J et al (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc 4(5):698–705

Chapter 21

From Phosphosites to Kinases

Stephanie Munk, Jan C. Refsgaard, Jesper V. Olsen, and Lars J. Jensen

Abstract

Kinases play a pivotal role in propagating the phosphorylation-mediated signaling networks in living cells. With the overwhelming quantities of phosphoproteomics data being generated, the number of identified phosphorylation sites (phosphosites) is ever increasing. Often, proteomics investigations aim to understand the global signaling modulation that takes place in different biological conditions investigated. For phosphoproteomics data, identifying the kinases central to mediating this response is key. This has prompted several efforts to catalogue the immense amounts of phosphorylation data and known or predicted kinases responsible for the modifications. However, barely 20 % of the known phosphosites are assigned to a kinase, initiating various bioinformatics efforts that attempt to predict the responsible kinases. These algorithms employ different approaches to predict kinase consensus sequence motifs, mostly based on large scale in vivo and in vitro experiments. The context of the kinase and the phosphorylated proteins in a biological system is equally important for predicting association between the enzymes and substrates, an aspect that is also being tackled with available bioinformatics tools. This chapter summarizes the use of the larger phosphorylation databases, and approaches that can be applied to predict kinases that phosphorylate individual sites or that are globally modulated in phosphoproteomics datasets.

Keywords Phosphoproteomics, Kinases, NetPhorest, NetworKIN, Phospho.ELM, PHOSIDA, PhoshoSitePlus

1 Introduction

Dynamic protein phosphorylation is a key regulatory mechanism by which cellular processes are initiated, repressed, and modulated to maintain homeostasis. The magnitude of the cellular phosphorylation network is reflected in a typical phosphoproteomics screen, which can identify thousands of phosphosites at a time [1, 2]. The timely and rigid control of the cell's elaborate phosphorylation signaling networks is achieved by the coordinated functions of kinases and phosphatases that, respectively, add and remove a phosphate group on serine, threonine, and tyrosine residues on proteins. The human proteome is estimated to contain well over 500 protein kinases that specifically phosphorylate their intended substrates to produce a desired and concise effect for maintenance of cellular

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_21,

functions [3]. A kinase's substrate specificity is in part conferred by preferences for certain amino acids or types of amino acids among the residues surrounding the target phosphosite (consensus motif). Equally important for the concise kinase mediated cellular signaling is the contextual activation and inactivation of the kinases, the regulation of its subcellular localization and that of its substrates. One interesting example are the ATM and ATR kinases that have the same known consensus phosphorylation motif, S/T-Q. It is therefore their timely activation in response to distinct perturbations, their localization, and the cellular distribution of their substrates that contribute to the regulation of their individual functions.

While phosphoproteomics is a powerful technology that allows deep and broad insight to the phosphorylation networks of a cellular system, the capacity to interpret the significance of these phosphosites is largely lagging. As such, the Phospho.ELM database currently comprises nearly 43,000 phosphosites, yet the function of most of these is unknown, and only a fraction of these phosphorylation events have been attributed to a specific kinase [4]. However, the central role of protein kinases in orchestrating the cell's signaling networks renders this family of proteins of particular interest in global investigations of perturbed and diseased systems, and not in the least for quantitative phosphoproteomics screens. The importance of clarifying the roles of kinases in biological systems is evident in that kinases are intensively investigated as potential drug targets, in particular for cancer therapy [5]. Evidently, a better understanding of the regulation of kinase activity and identification of its targets and their roles in a biological system will provide a foundation for further development of drugs for clinical use.

Phosphoproteomics data contain several levels of information that can be extracted or inferred to estimate the functional importance of individual kinases in a specific biological setting. Generally, the roles of kinases can be approached by analyzing the individual phosphosites or by determining which kinases are globally more active in the conditions investigated. For the individual phosphorylation events, a responsible kinase can be predicted based on the amino acids in the surrounding sequence. Furthermore, phosphorylation of kinases themselves often regulates their enzymatic activity. Therefore, a global analysis of which kinases appear to be active (or not) in the biological setting investigated can be deduced by bioinformatic analysis of complete phosphoproteomics dataset of phosphosites.

2 Phosphosites on Kinases

The most direct evidence of kinase activity modulation lies in the phosphosites identified on the kinases themselves. For some kinases, the exact phosphosites that control the activation of the enzymatic activity are known. If identified and quantified in a

phosphoproteomics screen, these sites provide the most direct indication of the induction or repression of a kinase's activity. Notably, the potential regulatory effects of most phosphosites on kinases are not known. Additionally, phosphoproteomics data does not reveal other aspects of kinase regulation, such as subcellular localization, as previously discussed. However, phosphorylation of activating sites in the kinase domain's activation loop can be used as a proxy for kinase activity. The activation loop of a eukaryotic protein kinase is contained within the activation segment in the kinase domain, which is located between a DFG motif and an APE motif. Introduction of a negatively charged phosphorylation site within this region activates most kinases by counteracting the positively charged arginine in the catalytic HRD motif, rendering the kinase domain in an active conformation.

2.1 Activation Loop Analysis Using UniProt Identifiers and Sites

At http://phomics.jensenlab.org we provide a tool to analyze phosphoproteomics datasets for phosphosites that reside within a kinase activation loop.

2.1.1 Protocol

1. Open the website: http://phomics.jensenlab.org
2. Click on "Activation Loop Analysis"
3. The phosphoproteomics data can be uploaded in one of the two ways:
   (a) Click on "Activation Loop Analysis Phosphosites" to upload either a tab delimited text file by clicking "Choose File" or copy-paste your sites into the "Input Sites" text area. Both should be in one of the following formats (*see* **Note 1**):
   - [uniProt ID] [tab or space] [amino acid] [Site]
   - [uniProt ID] [tab or space] [Site]
   - [uniProt ID] [tab or space] [Site] [tab or space] [amino acid]

   (b) Click on "Activation Loop Analysis Phosphopeptides," to upload either a tab delimited text file by clicking "Choose File" or copy-paste your sites into the "Input Sites" text area. Both should be in one of the following formats:
   - PEPtIDE
   - PEPpTIDE
   - PEPT(ph)IDE
   - Under "Digestion" indicate the maximum number of missed cleavages that were used in the database search, and the specificity of the enzyme used for sample preparation (*see* **Note 2**). Default settings are for trypsin with maximum two missed cleavages.
4. Click "Submit."

5. A results page will appear containing a table with the following four columns:
   (a) UniProt Entry: The UniProt accession number
   (b) Gene Names: The Gene name (as annotated in UniProt). Holding the cursor over "Gene Names" will give the UniProt protein names.
   (c) Loop Sites: All the phosphosites found within the kinase's activation loop sequence.
   (d) Other Sites: All other phosphosites found on the kinase.
   (e) If using phosphopeptide sequences as input, a column with the respective peptides will be listed in a separate column.
6. Click "Download Data" to download this table as a tab-separated file which can be opened in Microsoft Excel or Open/Libre Office.

3 Interpretation

To assess which protein kinases are differentially activated between conditions of interest, this analysis should be performed on comparative phosphosite datasets that have perturbed phosphorylation events in one condition compared to another. This will provide the best foundation for a biologically meaningful analysis, as the identification of regulated phosphosites in the kinase activation loop indicates that the activity of this kinase is possibly changed. This type of analysis can therefore be used to confirm or reject hypotheses about the kinases involved in cellular responses. It is however important to note, that this form of evidence is only indicative and it is important that the activity of the resulting kinases of interest, be validated by other biochemical assays. Equally important to keep in mind is the fact that the activation loop sequence is highly conserved and it is therefore often difficult to distinguish between closely related kinases. Additionally, the results of this analysis should be interpreted with caution, as there are other factors that determine the role of a kinase in the cellular signaling response networks. While the phosphorylation of the site may be regulated, the effect of the kinase in the system is not determined by this phosphorylation change alone. It can therefore recommendable to calculate the change in the stoichiometry of the phosphorylation, as the fold change does not indicate how much of the cellular pool of the given kinase is phosphorylated. However, this requires measurement of the proteome of the original sample in order to capture the unmodified peptides necessary for stoichiometry calculations [1]. If this data is not available, another option is to perform a sequence bias analysis of the regulated phosphosites compared to those that are not changed (*see* Chapter 23). Sequence bias can be matched to kinase phosphorylation consensus motifs

and can support findings of changes in the phosphosites of kinases in their activation loop. This analysis may mask the less regulated kinases. To gain knowledge of those kinases we recommend the two protocols below that employ NetworKIN to determine the kinases responsible for individual phosphorylation events and those that are generally regulated in the system.

4 Predicting Kinases for Individual Phosphorylation Events

Often, the identification of the kinases responsible for individual phosphorylation events is an essential factor for dissection of the biological question addressed with the phosphoproteomics screen. For known phosphosites, many curated databases provide information about kinases that are known to phosphorylate these sites, when such information is available from experimental data. These databases include PhosphoSitePlus [6], PHOSIDA [7, 8] and Phospho.ELM [4] (Table 1). The rapid advances in Phosphoproteomics have generated an explosion in the number of annotated phosphosites in these databases. For instance, in 2003 PhosphoSitePlus comprised 100 phosphosites mapped by mass spectrometry, in 2008 this number grew to 37,533, then 129,082 in 2012, and in 2014 consisted of 293,606 sites (*see* Table 1). It is therefore important when using these databases to be aware of how recently they have been updated. Similarly the coverage of different organisms must also be considered when choosing a database. The three aforementioned databases are briefly described below.

5 Databases

5.1 PhosphositePlus

PhosphoSitePlus currently covers of 176,152 phosphoserines, 68,622 phosphothreonines, and 48,832 phosphotyrosines (*see* Table 2), which are partially derived from published literature and in part unpublished in-house phosphoproteomics datasets generated by Cell Signaling Technologies, the company that curates the database. This renders PhosphoSitePlus the most

Table 1
Phosphosite databases

Database	Annotated phosphosites	Last update	URL
PhosphoSitePlus [6]	293,606	2014	http://www.phosphosite.org/
PHOSIDA [7, 8]	70,095	2012	http://www.phosida.com/
Phospho.ELM [4]	42,574	2011	http://phospho.elm.eu.org/

Table 2
Phosphosites in the PhosphoSitePlus database

	Phospho-serines		Phospho-threonines		Phospho-tyrosines	
Organism	**LT**	**HT**	**LT**	**HT**	**LT**	**HT**
Human	6515	101,493	2176	45,988	1970	33,739
Mouse	2300	55,978	719	16,793	829	8131
Rat	1476	7439	447	2245	413	3563
Other	632	319	149	105	146	41
Total	10,923	165,229	3491	65,131	3358	45,474

LT low-throughput, found using a 'site specific' method; *HT* high-throughput found using mass spectrometry

Table 3
Phosphosites in the Phospho.ELM database

	Phosphoserines		Phosphothreonines		Phosphotyrosines	
	LT	**HT**	**LT**	**HT**	**LT**	**HT**
Human	2113	17,316	637	4190	808	1586
Other	624	137	147	8	112	3
Mouse	598	4454	224	834	353	312
Worm	43	4453	27	987	9	102
Fly	66	2168	16	450	25	59
Total	3444	28,528	1051	6469	1307	2062

LT low throughput, found using a "site-specific" method; *HT* high throughput found using mass spectrometry

extensive phoshosite database to date. This resource covers an impressive number of human, mouse and rat phosphosites, while for other organisms the user may be better served with alternative databases such as Phospho.ELM (*see* Tables 2 and 3). The web interface allows the user to query a protein of interest, for which a result page including a detailed description of the proteins function, cellular localization, molecular function and other types of information is provided. All known annotated modified sites on a protein (including phosphorylation, acetylation, SUMOylation, and ubiquitination) are provided as a list. These are also visualized on a schematic drawing of the protein that contains additional information about the protein's domains. The user can seek further information for each site by clicking on it, upon which all the evidence, including references and potential enzymes responsible for the modification, can be retrieved.

6 Phospho.ELM

Phospho.ELM is a database covering phosphosites identified from in vivo or in vitro phosphoproteomics studies and from literature [4]. This database currently consists of 42,573 unique phosphosites, 31,754 of which are phosphoserine, 7449 are phosphothreonine, and 3370 are phosphotyrosine (*see* Table 3). Upon querying a protein, the web-interface provides a useful header, with links to the results from other resources such as the STRING database, NetworKIN and PHOSIDA. Additionally, a list of the annotated phosphosites is provided along with the surrounding amino acid sequence. If a kinase is annotated to phosphorylate the given site it is listed along with the evidence (reference) justifying this association. Other potentially useful information for phosphoproteomics investigations includes a conservation score and a surface accessibility score.

7 PHOSIDA

The incentive behind the PHOSIDA (PHOsphorylation SIte DAtabase) database was to generate a publicly and readily accessible repository of phosphorylation data. This effort was initiated on the basis of a seminal study published by the Mann laboratory, in which they conducted a SILAC-based time-resolved phosphoproteomics study of EGF (epidermal growth factor)-stimulated human cells [9]. The database has since developed to include other high-mass-accuracy, mass spectrometric phosphoproteomics data sets and other modifications, such as acetylation and N-glycosylation—all of which are from the Mann laboratory. PHOSIDA is therefore now known as the 'Posttranslational Modification Site Database'. In addition to human data, this resource also stores modification site information for mouse, fly, worm, and yeast.

8 High-Throughput Prediction Approaches

Evidently, many efforts have been initiated to organize and provide information on known phosphosites. However, with the ever-increasing number of mapped phosphorylation events, the effort to identify which of the approximately 518 protein kinases that are responsible for individual phosphorylation events is lagging behind. As such, of the nearly 43,000 phosphosites that have been curated and annotated in the Phospho.ELM database, only approximately 20 % have been linked to a kinase [4]. Additionally, for large phosphoproteomics datasets, this manual lookup approach is very cumbersome. Together these challenges have driven the development

of computational tools that apply algorithms to predict which kinases are responsible for phosphorylating individual sites.

As kinases are known to have preferences for linear motifs, which they recognize as a target for phosphorylation, using these consensus motifs to predict kinase-target associations would seem intuitive. Yet, in most cases it is very difficult to identify a linear amino acid motif that a kinase recognizes, as the number of known target sites for most kinases is not sufficient to produce a statistically sound phosphorylation consensus motif [10]. In principle this obstacle can be approached experimentally by using synthetic peptide libraries, and approach that has been used to identify the linear motifs applied in the Scansite web tool [11, 12] (http://scansite.mit.edu/). Here a degenerate peptide library, with a centered phosphorylation was exposed to a particular kinase, and the resulting phosphorylated peptides were collectively sequenced by Edman degradation [13]. This revealed the relative amount of each amino acid at a given position, which, when used to produce position-specific scoring matrices (PSSMs), could reveal statistically significant preferences of the kinase for each amino acid type in the sequence surrounding the phosphosite. The Scansite web-based tool allows the user to search protein sequences for phosphorylation motifs recognized by kinases.

Similarly, NetPhorest (http://netphorest.info/) provides a comprehensive atlas of linear motifs recognized by specific kinases [10]. NetPhorest maps existing in vitro and in vivo datasets of protein phosphosites, which are linked to at least one kinase, onto the phylogenic tree of the kinases. As such, a redundancy-reduced positive and negative control datasets can be produced for each kinase or family of kinases. With this basis, linear motif classifiers are predicted using artificial neural networks (ANNs) and PSSMs (as with Scansite). The use of ANNs is beneficial as they detect nonlinear correlations between residues, and these are trained on in vivo and peptide-based in vitro phosphorylation data for a specific kinase or kinase family. PSSMs are constructed using information from other in vitro based approaches for the specific kinases. A probabilistic score is calculated and employed to determine the best classifiers for a kinase or group of kinases after several levels of filtering and benchmarking. The authors behind this effort found that it is unlikely to determine a consensus motif classifier for a specific kinase, but that motifs can be more reliably predicted for families of kinases [10].

While these sequence-based kinase-substrate associations are very useful, kinase specificity is also conferred by context, as previously described. This aspect of the kinase-substrate interaction is not accounted for when using consensus motifs to predict the kinase responsible for individual phosphorylation events. The effort

to account for this functional association is the unique strength of the NetworKIN algorithm (http://networkin.info/) [14, 15]. For each queried phosphosite, this Web-based tool combines the consensus sequence classifiers from NetPhorest with the network context for the kinases connected to the substrate from the STRING database [16]. To prioritize the "best-fitting" kinase, this algorithm scores the results from the two databases, and combines them, resulting in more accurate prediction of kinase-target association. As such, a NetPhorest probability is calculated for the kinase families from the amino acid sequence surrounding the queried phosphosites using the NetPhorest classifiers. Similarly, for each queried protein a network proximity score is calculated for all kinases, by multiplying the confidence score for each edge necessary to connect the kinase and the substrate [14]. The NetPhorest and network probabilities are each converted to "likelihood scores," and the network likelihood scores are further corrected for biases derived from over-studied proteins (which are inherently overrepresented in the databases). These two likelihoods for each kinase are combined to produce a "unified likelihood ratio." The NetworKIN output thereby provides a list of kinases predicted to be responsible for phosphorylating the queried phosphosite, with the unified likelihood ratio for each potential kinase provided as an indicator of the confidence of the prediction, and where a higher score indicates a more confident prediction. The stringency of the score cutoff for the kinases listed in the results interface can be manipulated to provide a list of more or less confident hits. Furthermore, the underlying evidence from both the STRING database and NetPhorest can be extracted, such that the user can more carefully examine the basis for two closely scoring kinases. However it must be noted that NetworKIN is only applicable for human and yeast data.

9 Protocol: NetworKIN—Predict Enzymes for Observed Phosphosites

9.1 Material

1. Datasets. A list of phosphosites of interest. These should be tab or space delimited, with three columns: (1) protein IDs (for example, UniProt, Ensembl, or RefSeq); (2) position of phosphorylation in the protein; and (3) residue (serine, threonine, or tyrosine). Alternatively the tab-delimited MaxQuant [17] output Phospho(STY) table can also be directly uploaded.
2. Software. This protocol is conducted on the Web-based high-throughput version of NetworKIN available at http://networkin.info/index_ht.shtml. The query datasets are deleted after processing. If speed or confidentiality is of high priority a stand-alone version is also available (*see* **Note 3**).

10 Methods

1. Upload Data
   Select an appropriate species and database in the "Select the sequence database" drop-down menu. Upload the Phospho(STY).txt file from MaxQuant or paste the list of protein identifiers, position, and residues (tab or space delimited) into the designated space. You may be asked to disambiguate your identifiers if they do not map uniquely to the STRING identifiers. The tool will process your data. Depending on the size of your dataset the processing can take some time.
2. *NetworKIN output*
   The NetworKIN output is visualized in a results interface (*see* Fig. 1). Here the user can change the default settings for minimum score and the maximum difference accepted (*see* **Note 4**). In addition to predicting kinases, NetworKIN also predicts possible phospho-binding domains on proteins that may associate with the input phosphosites. Disable this function.

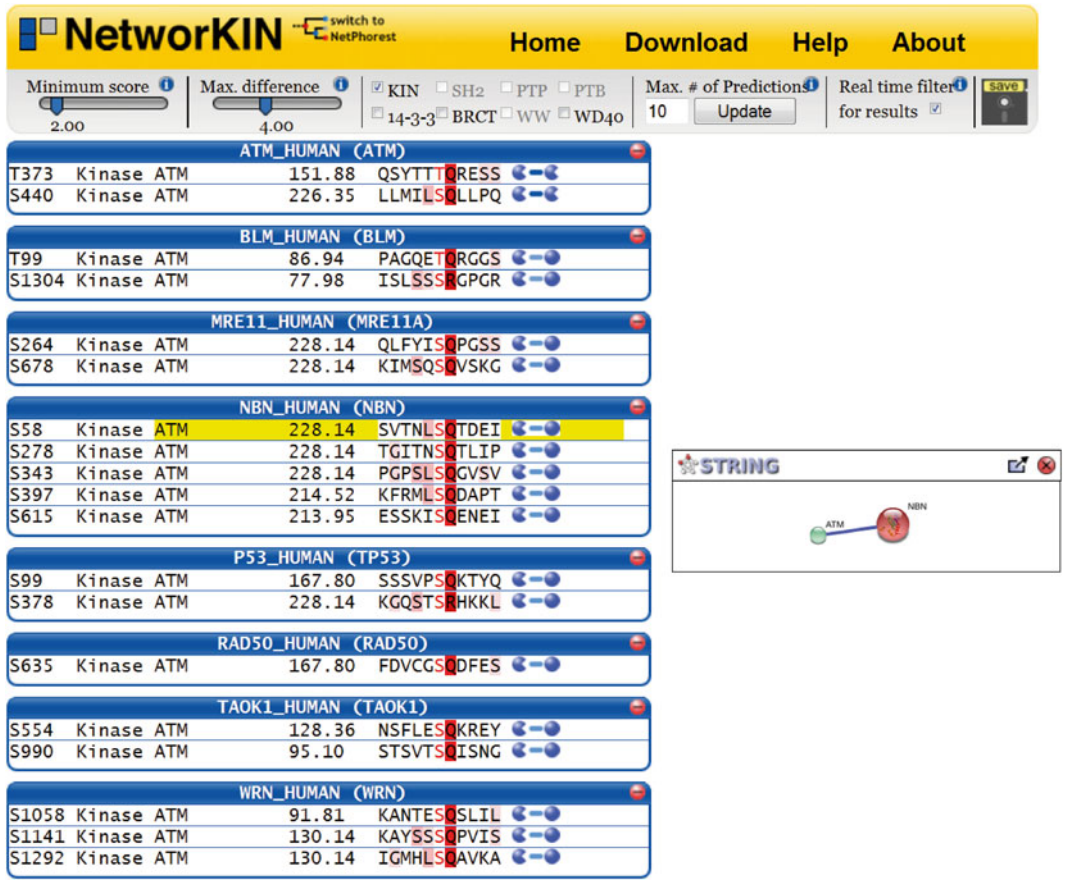


Fig. 1 NetworKIN. The web interface for inspecting results from NetworKIN

The results interface lists every input protein identifier within a blue box, with each site separated. Each line herein is dedicated to a predicted enzyme, and the 'unified likelihood ratio' is reported beside it. For each prediction, the NetPhorest classifier motif is highlighted in a sequence window surrounding the phosphosite, where the saturation of red indicates the confidence of the motif as a predictor for the kinase. Additionally, a schematic representation of a network indicates how many nodes and edges are required to link the kinase and the substrate. The kinase is the incomplete circular node, the substrate is the blue node, and grey shapes indicate intermediate nodes. By clicking on this, the user is provided with the STRING network evidence (*see* Fig. 1).

3. *Exporting*
   The data can be exported by clicking the diskette icon, which allows the user to download the full dataset (all the information for all predicted kinases) or the filtered dataset (presented on the interface).
4. *Interpretation*
   As previously described, the NetworKIN analysis provides a list of the kinases that are predicted to phosphorylate each of the queried sites. The kinases displayed upon analysis are those that meet the criteria for the default "Minimum score" and "Maximum difference" settings. However, it is again important to keep in mind that these are merely predictions, and it is recommendable to examine the results in more detail to verify that the evidence is biologically meaningful, particularly if the results are intended for follow-up experiments. In most cases the user can chose to accept the highest scoring kinases for further analysis. However, in some cases, there may be two protein kinases with very similar scores, in which case the user will have to make a judgment call based on the underlying evidence. While it is recommendable to apply stringent filters for most phosphoproteomics data analysis, it can be beneficial to lower the "Minimum score" and "Maximum difference" cutoffs in NetworKIN analysis. As such, less confident candidate kinases will be displayed, and in rare cases these can be worth considering for further analysis, based on evaluation of the evidence.

11 Predicting Kinases Regulated in a Biological System

NetworKIN provides a powerful tool to predict kinases that are responsible for individual phosphorylation events. However, often a phosphoproteomics investigation prompts the question of which kinases are generally activated or deregulated under the conditions being tested. Such analyses necessitate that the phosphoproteomics

experiment is performed quantitatively, such that phosphosites that are altered in the biological conditions can be distinguished from those that are unperturbed (*see* Chapter 22). As is the case for most bioinformatics analysis, we recommend that in addition to generating a dataset of phosphosites of interest, a reference dataset consisting of phosphosites that are unchanged be extracted from the same experimental data. With this approach, biases in sample preparation and MS data acquisition are inherently accounted for, preventing inaccurate interpretations of the analysis results.

The most commonly used approach to infer which kinases have altered activity in an experimental setup, is to generate a sequence logo plot. These are readily generated in many web-based tools, some of which allow for foreground and background input datasets (*see* Chapter 23). On the nominator, the sequence logo plot indicates which amino acids are overrepresented among the residues surrounding the phosphosite in the foreground data compared to the background. Similarly, the denominator indicates underrepresented amino acids. The resulting motifs can be compared to known kinase family consensus phosphorylation motifs to predict their modulated activity in the phosphoproteomics study.

This method is however only indicative and not very robust. We therefore recommend an alternative approach, in which NetworKIN is used to predict kinases responsible for each phosphorylation event in the foreground and background datasets, such that their prevalence can be compared between these two datasets. Furthermore, a Kolmogorov–Smirnov test can be performed between two NetworKIN output files, to estimate which kinases are significantly enriched.

12 Protocol: Identifying Globally Enriched Kinases

12.1 Materials

1. Datasets: Two datasets containing phosphosites (*see* Subheading 3.3.1 for input data format). The two datasets should consist of one foreground dataset with phosphosites of interest (foreground) and a reference dataset (background). *See* Chapter 22 for a guide to define these.
2. Software: http://networkin.info/index_ht.shtml and http://phomics.jensenlab.org/kinase_enrichment

13 Methods

1. For each dataset, follow all steps of protocol 3.3 "Predict enzymes for phosphosites" and export the "full dataset."
2. Go to http://phomics.jensenlab.org/kinase_enrichment and upload these NetworKIN output files.
3. Click download.

14 Interpretation

The results are exported as a tab-separated file that can be opened in Excel. Herein, there is one row for each kinase enriched in one dataset (foreground) compared to the other (background). The "Entry" column lists the highest scoring kinase from the respective family. The "P-value" column shows the p-value for the two sided Kolmogorov–Smirnov test, and "d-value" indicates the distance measure from the Kolmogorov–Smirnov test (where a positive value signifies that it is more prevalent in the foreground compared to the background). Because the NetworKIN likelihood ratios stem from motif and context scores, and the motif score is often the same for all kinases in a family, this method can often not dissect exactly which kinase is active. This furthermore makes it difficult to correct for multiple testing, as the tests are not independent. It is therefore most meaningful to analyze the protein kinase families that are enriched, and for further experiments to use existing knowledge to predict the kinase(s) in the family that may be most relevant.

As is the case with most bioinformatics analysis, it is important that all results be treated with caution, as they are predictions. The user's biological knowledge is ultimately the decisive factor for meaningful interpretation and prioritization for functional validation of the predictions. The user must therefore investigate the biological significance, for example by checking whether the kinase and the substrate are expressed in the same tissue. This contextual information is in part included in NetworKIN due to the use of STRING data, but must be further elaborated.

15 Conclusion

While there is no definitive *in silico* means to predict kinases that phosphorylate proteins, efforts in recent years have provided tools that give a good foundation for interpretation of phosphoproteomics data. Here we presented three types of analysis, and have listed several useful databases. It is ultimately the aim of the investigation that determines the most appropriate approach(s) that will provide the necessary information. High-throughput tools, such as NetworKIN, are particularly advantageous for large phosphoproteomics datasets, and can with great benefit be followed up by closer examination of information stored in the many databases available. The results from most of these analyses are essentially predictions, and must be tested experimentally to draw definitive conclusions.

16 Notes

1. To view examples of the valid input formats, click "Example data" and the examples will appear in the "Input Sites" text area. The three first examples demonstrate the three compatible formats.
2. Most phosphoproteomics pipelines require that the proteins from the original sample are digested with proteases to produce peptides for analysis by LC-MS/MS.
3. Download at http://networkin.info/download/NetworKIN3.0_release.zip and follow the readme.
4. With "Minimum score" the user can define the lower threshold for the Unified Likelihood ratio for the kinases appearing the results page. The "Maximum difference" setting defines what the limit of inclusion for kinases with confidence below that of the best scoring enzyme.

Acknowledgments

This work was in part funded by the Novo Nordisk Foundation Center for Protein Research [NNF14CC0001].

References

1. Olsen JV et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3(104):ra3
2. Sharma K et al (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep 8(5):1583–1594
3. Manning G et al (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934
4. Dinkel H et al (2011) Phospho.ELM a database of phosphorylation sites--update 2011. Nucleic Acids Res 39(Database issue):D261–D267
5. Rask-Andersen M et al (2014) Advances in kinase targeting: current clinical use and clinical trials. Trends Pharmacol Sci 35(11): 604–620
6. Hornbeck PV et al (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 40(Database issue):D261–D270
7. Gnad F, Gunawardena J, Mann M (2011) PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 39(Database issue):D253–D260
8. Gnad F et al (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol 8(11):R250
9. Olsen JV et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635–648
10. Miller ML et al (2008) Linear motif atlas for phosphorylation-dependent signaling. Sci Signal 1(35):2
11. Yaffe MB et al (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat Biotechnol 19(4): 348–353
12. Obenauer JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 31(13):3635–3641
13. Songyang Z et al (1994) Use of an oriented peptide library to determine the optimal

substrates of protein kinases. Curr Biol 4(11): 973–982

14. Horn H et al (2014) KinomeXplorer: an integrated platform for kinome biology studies. Nat Methods 11(6):603–604
15. Linding R et al (2007) Systematic discovery of in vivo phosphorylation networks. Cell 129(7):1415–1426
16. Franceschini A et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41(Database issue):D808–D815
17. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372

Chapter 22

Search Databases and Statistics: Pitfalls and Best Practices in Phosphoproteomics

Jan C. Refsgaard, Stephanie Munk, and Lars J. Jensen

Abstract

Advances in mass spectrometric instrumentation in the past 15 years have resulted in an explosion in the raw data yield from typical phosphoproteomics workflows. This poses the challenge of confidently identifying peptide sequences, localizing phosphosites to proteins and quantifying these from the vast amounts of raw data. This task is tackled by computational tools implementing algorithms that match the experimental data to databases, providing the user with lists for downstream analysis. Several platforms for such automated interpretation of mass spectrometric data have been developed, each having strengths and weaknesses that must be considered for the individual needs. These are reviewed in this chapter. Equally critical for generating highly confident output datasets is the application of sound statistical criteria to limit the inclusion of incorrect peptide identifications from database searches. Additionally, careful filtering and use of appropriate statistical tests on the output datasets affects the quality of all downstream analyses and interpretation of the data. Our considerations and general practices on these aspects of phosphoproteomics data processing are presented here.

Key words Phosphoproteomics, Database Search, False Discovery Rate, Statistics, Quantitation, MaxQuant

1 Introduction

Virtually all cellular processes are regulated by posttranslational modifications (PTMs). Phosphorylation is a crucial and highly dynamic PTM that contributes to cellular physiology and pathophysiological developments [1]. Phosphoproteomics platforms are generating an ever increasing amount of data. For the experimentalist the challenge lies in transforming these vast amounts of information in the acquired MS and MS/MS scans into quantified phosphorylation sites (phosphosites) mapped to identified proteins.

There are several approaches for assigning peptide sequences to ions sequenced by mass spectrometry (MS). De novo sequencing reads out the peptide sequence from the mass differences of the ions detected in the MS/MS scan [2, 3]. In the early years of the mass

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_22, © Springer Science+Business Media New York 2016

spectrometry era, this daunting task was performed manually, but it is now automated. An alternative to de novo approaches is the spectral library approach, which compares each MS/MS spectrum to a reference library of previously observed MS/MS spectra [4]. Another approach is the database search strategy, which implements theoretical spectra from an in-silico digested database of all proteins in the species of interest [5]. The acquired experimental spectra can be compared to those theoretical spectra to infer and score peptide spectral matches (PSMs). The latter approach is the most widely used in phosphoproteomics analyses. This strategy has also been extended to mapping and scoring of phosphosites. As such, for every potentially phosphorylated peptide, a theoretical tandem spectrum is generated for each possibly phosphorylated version of the peptide (corresponding to each serine, threonine and tyrosine in the peptide). The platforms of phosphorylation site identification reviewed in this chapter implement this latter strategy.

There are three independent steps in the processing of raw MS data into quantified phosphosite ratios:

1. *Database search:* Raw spectra are searched against a reference peptide database, and a score is calculated for each PSM.
2. *Filtering:* In this step the PSMs are sorted and filtered down to a target False Discovery Rate (FDR) to limit false positive identifications.
3. *Quantitation:* Finally ratios are calculated for all peptides (and proteins).

This chapter gives an overview of all the important considerations which should be taken into account when processing raw data to retrieve a quantified phosphoproteomics output. The focus is on understanding the biases that are inherent to such data so that common pitfalls can be avoided.

2 MS Data Formats

Most vendors of mass spectrometric instrumentation have their own MS raw file format, and generally also provide a platform to process this raw data into quantified data. However, if the user opts for software that is unable to parse these formats, conversion will be necessary. Phosphoproteomics data generally contains two types of information: (1) full MS spectra and (2) MS/MS (tandem MS) spectra. When converting from vendor raw data format, it is important to bear in mind the information retained in the new file format. As such, the popular format MGF (Mascot Generic File) does not contain full MS information, and it therefore cannot be used for quantification based on metabolic labeling [6, 7]. For converting between MS file formats we recommend ProteoWizard MSConvert [8] or TOPPAS [9] FileConverter workflow.

ProteoWizard MSConvert [8] is available for Windows, Linux, and OS X. All versions of the software can read and write the following open formats: mzML, mzXML, MGF, ms2/cms2/bms2, and mzIdentML. The Windows version can furthermore read the following vendor formats: Agilent, Bruker FID/YEP/BAF, Thermo RAW, and Waters RAW. MSConvert has both a command line and graphical user interface, rendering it versatile and user friendly.

TOPPAS FileConverter [9] is available for Windows, Linux, and OS X, and it provides a graphical user interface to the command line tool FileConverter, which is part of TOPP (The OpenMS Proteomics Pipeline) [10, 11]. It can convert the following input file formats: mzData, mzXML, mzML, dta, dta2d, MGF, featureXML, consensusXML, ms2, fid, tsv, peplist, kroenik, and edta into mzData, mzXML, mzML, dta2d, MGF, featureXML, consensusXML, and edta format.

In our opinion the best open formats are mzML and mzXML. mzML is the de facto standard that was developed by the HUPO (Human Proteome Organization) initiative to unify the mzXML and mzData formats.

For more in-depth overview of all proteomics file formats the reader is referred to Deutsch [12].

3 Database Search

A wealth of different database search algorithms has been developed over the years. The most popular include the open source engines X!Tandem [13], OMSSA [14], MyriMatch [15], the freeware engine Andromeda [16], and the proprietary engines SEQUEST [17], PEAKS DB [3], and MASCOT [18]. All these algorithms are very mature and should produce comparable results. However, because they all use slightly different and to some extent orthogonal scoring schemes, using two in combination often yield higher confidence identifications [19].

4 Filtering

This step of the phosphosite identification workflow aims to exclude low-confidence identifications resulting from the database search. While more stringent filtering will result in a sacrifice of identifications of the total number of peptides and phosphosites, the resulting identifications are more trustworthy. This is in particular advantageous for the experimentalists using this information for hypothesis generation and downstream experiments.

There are two commonly used approaches to filter MS data, based on (1) arbitrary PSM score cutoffs and (2) FDR. The PSM score cutoff strategy depends on the size of the search database.

Larger databases will inherently yield more false positive identifications at a given cutoff. Therefore higher PSM score cutoffs are required when dealing with phosphoproteomics in comparison to proteomes due to the differences in search space. Therefore we recommend the FDR approach.

4.1 False Discovery Rate

The advantage of the FDR approach over the cutoff approach is that a given FDR is comparable across datasets, irrespective of the size of the dataset and search space. FDR is calculated using a target–decoy database approach, in which the PSMs are performed against a database of interest and a fictive database of comparable size. The most widely used decoy approach is the pseudo-reversed method proposed by Elias and Gygi [19], in which they dubbed the decoy database "reverse" and the original search database the "forward." Pseudo-reversed is a quite fitting name, as the tryptically digested peptides are literally reversed, except for the last R/K, which is swapped. Searching against a concatenated forward and reverse database, the FDR can be calculated simply by counting the number of forward and reverse peptides/proteins above a given score cutoff. For most search engine platforms, the desired FDR is set in advance, and all peptides above the corresponding score are accepted. Additionally, most platforms allow the user to set the FDR on both peptide and protein identifications. Setting the FDR on the protein identification level will almost always result in a more stringent FDR of the peptide identifications, as multiple forward hits match the same protein. For phosphoproteomics data, however, it is recommendable to set the FDR on the peptide level, as the resulting phosphosites are identified on the peptide level and ultimately the data of interest. A 1 % FDR is commonly accepted in the phosphoproteomics community.

4.2 Phosphosite Localization Probability

The ultimate aim of phosphoproteomics investigations is to identify the exact location of the phosphorylation moiety in the sequenced phosphopeptide, and the predicted protein of origin. While no successful approach has been developed for implementing false localization rates at the phosphorylation site level, applying stringent FDR at the peptide level is beneficial in the processing of phosphoproteomics data. Peptides that pass at a higher FDR cutoff are generally identified from MS/MS scans with more peaks, thereby increasing the confidence of the phosphorylation site localization. This is illustrated in Fig. 1 where the median localization probability drops below 100 % at peptide FDR = 0.2 % (*see* Fig. 1). In addition to a stringent FDR at the peptide level, it is common practice to filter all resulting phosphosites such that none has a localization probability below 75 % [20]. While 75 % may seem very low, it is important to note that most sites can be localized with much higher accuracy, some even with 100 %

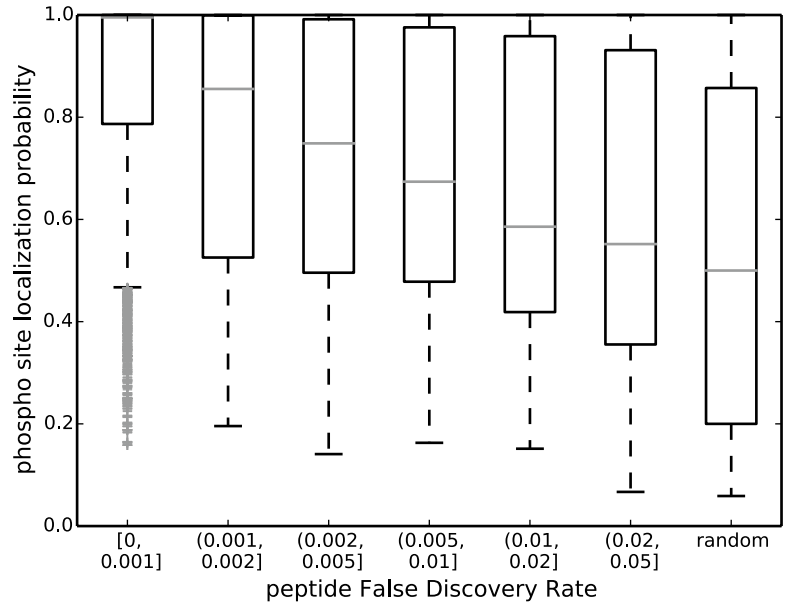


Fig. 1 Boxplots depicting the relationship between the phosphorylation localization probability and the False Discovery rate

accuracy if the peptide contains only one serine, threonine or tyrosine. The average localization probability at 1 % peptide FDR was 96.6 % for the data used to create Fig. 1.

5 Quantitation

Phosphoproteomics experiments often seek to determine the differences in the abundance of phosphosites between perturbations, tissues or other conditions, as these changes are vital to answering the biological question at hand. Quantitation strategies are either based on label-free approaches that do not require alterations to the experimental workflow, or on isotopic labeling of amino acids. Label-free quantitation (LFQ) is based on comparison of full MS scans from separate MS raw files representing the conditions of interest. Recent innovations in LFQ such as iBAQ [21] combined with improvements of MS instrumentations have rendered LFQ strategies a viable alternative to isotopic labeling. While LFQ gives hopes for clinical proteomics, labeling-based strategies remain the preferred tool in quantitative phosphoproteomics. The two most popular labeling techniques are stable isotope labeling with amino acids in cell culture (SILAC) [6, 7] and isobaric tags for relative and absolute quantitation (iTRAQ) [22]. The advantage with labeling strategies lies in the detection of all conditions within the same MS or MS/MS scan, allowing for more direct and accurate quantitation.

6 Foreground and Background for Enrichment Analyses

The goal of most phosphoproteomics experiments is to understand how phosphorylation events affect a biological system. Evidently, phosphosites that differ between biological conditions are of interest, and such phosphoproteomics experiments should ideally be performed quantitatively. These phosphosites of interest are best understood in the context of the phosphosites that are unchanged in the given experimental data. When possible, it is therefore advisable to generate a reference dataset of unchanged phosphosites (background) and a dataset of changed phosphosites (foreground) from the same data as both will have been subjected to the same experimental biases. These datasets are typically defined by applying statistical tests, such as *t*-tests, to determine the fold change cutoffs.

6.1 Sources of Experimental Biases

Various steps in the phosphoproteomics workflows will inherently introduce biases in the data that can be misinterpreted as biologically relevant if not accounted for. The easiest and most robust approach that we recommend, as previously mentioned, is to generate a reference dataset from the same original data. Common sources of bias include the lysis buffer used, fractionation methods, phosphopeptide enrichment protocols, and MS methods applied. Phosphopeptide enrichment protocols commonly display biases towards either singly or multiply phosphorylated peptides or towards hydrophobic or hydrophilic peptides. Mass spectrometric biases are specific to the fragmentation technique and/or instrument applied. In MS methods where the most intense peaks in a MS spectrum are submitted for MS/MS analysis, there is an inherent bias to sequence peptides that are more abundant. The combination of all the abovementioned biases will be present in the experimental data, stressing the importance of applying a custom reference dataset.

6.2 Example Pitfall Caused by Experimental Bias

The following example systematically dissects a fictive dataset, to illustrate the importance of using an appropriate foreground and background. For simplicity, it is assumed that every protein only gives rise to one phosphorylated peptide.

6.2.1 Example Experiment

Two experimental conditions: control and perturbation.

Our fictive organism has 40,000 different proteins, of which 100 are ribosomal.

4000 SILAC phosphopeptide pairs are identified, of which 50 are ribosomal.

400 of them have a perturbation:control ratio above 5, of which 5 are ribosomal.

The aim is to calculate whether this perturbation of interest enriches ribosomal phosphosites. Using the above data we explore the importance of using a custom background.

6.2.2 Faulty Approach: Global Background

It is common in proteomics data analysis to compare an experimental foreground dataset to an entire proteome database of interest to search for significantly perturbed pathways, sequence motifs, and more. However, this approach assumes that the preparation and analysis of the sample does not introduce any biases. With this approach, when calculating the enrichment factor of ribosomal protein phosphorylation in your data compared to a complete proteome for this fictive dataset (n denotes "number of"):

Expectation: $n_{\text{ribosomal proteins}}/n_{\text{total proteins}} = 100/40{,}000 = 0.25\ \%$

Observation: $n_{\text{ribosomal proteins above cutoff}}/n_{\text{total proteins above cutoff}} = 5/400 = 1.25\ \%$

Enrichment: Observation/Expectation = 1.25 %/0.25 % = 5

Significance: A two-sided binomial test with 400 trials, 5 successes and an expected frequency of 0.25 % gives a *p*-value of 3.6 %, and the enrichment would thus be deemed significant.

Conclusion: The applied experimental perturbation increases phosphorylation on ribosomal proteins fivefold.

However, our fictive dataset has a bias, which is common to most phosphoprotemics workflows, namely an enrichment for abundant proteins. Therefore, 1.25 % of the identified phosphosites reside in ribosomal proteins, compared to only 0.25 % in the total proteome. We therefore encourage using the unperturbed experimental data itself as background when applying this custom background from the experimental data:

Expectation: $n_{\text{ribosomal proteins}}/n_{\text{total proteins}} = 50/4000 = 1.25\ \%$

Observation: $n_{\text{ribosomal proteins above cutoff}}/n_{\text{total proteins above cutoff}} = 5/400 = 1.25\ \%$

Enrichment: Observation/Expectation = 1.25 %/1.25 % = 1

Significance: A Fisher's exact test using a 2×2 contingency table with the values (5, 400) vs. (50, 4000) yields a *p*-value of 100 %, meaning there is a 100 % chance there is no difference between the number of regulated ribosomal proteins and other regulated proteins.

Conclusion: The applied experimental perturbation does not affect phosphorylation on ribosomal proteins.

6.3 Statistical Test Used to Define Foreground and Background Datasets

Foreground and background datasets are typically defined using statistical tests to determine appropriate cutoffs to judge whether a phosphosite is changed or unchanged between experimental conditions. Here we discuss common statistical approaches that can be applied to phosphoproteomics data and when those different tests are appropriate to apply.

Many statistical tests assume that the data conforms to the normal distribution and that the data has the same mean and variance across datasets in experiments with many conditions. It is therefore

important when analyzing quantitative data, to perform all statistical tests on log transformed data (usually $\log_2$ or $\log_{10}$), as fold change ratios in a linear scale are not normally distributed. Most proteomics software packages also normalize the output data to compensate for this.

Cutoffs are usually set using *p*-values, the motivation for which is to ensure that the entries in the foreground are significantly different from the background. However, when searching for biologically meaningful regulatory events, it is important to bear in mind that *p*-values only define whether an entity is significantly different from another. Thus, *p*-values are very susceptible to the number of replica and sample variance, which in the correct combination can lead to very small fold change being considered significant.

The nature of the phosphoproteomics data will determine the ideal statistical test, which should be used to calculate *p*-values. In cases where many replicate experiments have been performed, a student's *t*-test is advisable. This test requires many data points (at least three ratios) for each phosphorylation site, as this test is based on both the fold changes and the variance of each ratio across measurements. As such, the student's *t*-test also compensates for experimental and instrumental imposed variance, and should be used when those are expected. However, most phosphoproteomics screens are not performed with enough replicate experiments to perform student's *t*-test, in which case, a test based on the distribution of the entire dataset is beneficial. For this we recommend the Significance A test.

6.3.1 Student's t-tests

There are two types of *t*-tests that in principle could be applied to phosphoproteomics data: the related and the independent *t*-test. Both tests determine whether a phosphorylation site differs between conditions or not, and both are performed on the abundance values of the phosphosites rather than their ratios.

Related t-test: This test compares the abundance of phosphosites between conditions within each replicate experiment. This means that the abundances in condition 1 and 2 are compared within the first replicate, then within the second replicate and so forth. In theory this test has good statistical power, but it is limited by the fact that MS data often has many missing values, in which case the replicate data-points are lost, and the test loses its power.

Independent t-test: The independent *t*-test determines the statistical difference between the mean intensities of a phosphosite across conditions. As this test uses the mean intensities, missing values are tolerable, and it is therefore more robust for MS data compared to the related *t*-test.

Both of these *t*-tests can be performed assuming either equal or unequal variance. Equal variance calculates a variance of the abundance of the phosphosite across conditions and replicates, and the test will therefore use all data points. Unequal variance assumes

that the variance is not comparable between conditions, and the test will therefore calculate a variance across replicates but separately for each condition. This will therefore include less data points per variance calculation, and as missing values are common in MS data, applying equal variance is more robust.

6.3.2 Significance A

Significance A is a statistical test that determines for each ratio whether it differs significantly from the distribution of the whole dataset [23]. This test is particularly applicable for phosphoproteomics data that is heavily perturbed. This test takes advantage of the fact that the middle 68.26 % of the ratio distribution conforms better to a normal distribution.

Here the 15.87 %, 50 % and 84.13 % percentiles are assumed to correspond to left standard deviation r_{-1}, the mean r_0 and the right standard deviation r_1 respectively. The distance z (analogous to the standard deviation in t-test) is then calculated and applied to determine a p-value for the ratio of each phosphorylation site:

$$\text{if } r > r_0 \text{ then } : z = \frac{r - r_0}{r_1 - r_0}$$

$$\text{if } r < r_0 \text{ then } : z = \frac{r - r_0}{r_{-1} - r_0}$$

$$\text{Significance A} = \frac{1}{2} erfc\left(\frac{z}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} \int_z^{\infty} e^{-0.5t^2} dt$$

In the publication introducing Significance A, the authors also proposed the option of correcting the p-values for multiple testing. This can be done with a Benjamini–Hochberg correction.

To get an intuitive idea of how Significance A works, let us imagine data ratios that are distributed as follows:

9000 peptides are unregulated; average log ratio = 0 and a standard deviation = 1

1000 peptides are regulated; average log ratio = 0 and standard deviation = 5

Table 1 outlines the difference between using the "real" standard deviation (which we know for the used model dataset) and using Significance A or RMSD (Root Mean Square Deviation) to calculate the standard deviation (*see* Table 1). First it should be noted, for the above data, significance A overestimates the standard deviation by 21 % whereas RMSD overestimates it by 84 %. Using RMSD only 269 of the 1000 regulated proteins would be found above 3 standard deviations (used as cutoff) as estimated by RMSD. There against 469 proteins would pass this criterion when using significance A, which is much closer to what we would have gotten had we known the "real" standard deviation (549).

Table 1
Table showing how Significance A compares to root mean square distance (RMSD) when estimating standard deviation (SD) on a set of peptides where 9000 are unregulated and have a average log ratio of 0 and a SD of 1 and 1000 are regulated and have a average log ratio of 0 and a SD of 5

	Real	Significance A	RMSD
Standard deviation	1.00	1.21	1.84
Regulated peptides (of 1000) at 3 SD	549 (54.9 %)	469 (46.9 %)	269 (26.9 %)
Unregulated peptides (of 9000) at 3 SD	24 (0.270 %)	3 (0.029 %)	0 (3.2×10^{-6}%)

6.4 Approaches to Applying Cutoffs to Define Foreground and Background Dataset

There are several common approaches to define foreground or background datasets. These typically apply cutoffs for fold change, *p*-values or a combination thereof based on the quantitative phosphoproteome or the proteome from the same experiment. Choosing a good cutoff requires several considerations that we review here.

6.4.1 Setting Cutoffs Based on the Experimental Proteome vs. Phosphoproteome

If the proteome in the given data is expected to be largely unperturbed, for example in short term treatment conditions, the distribution of the proteome can be used to define a cutoff for the phosphoproteomics data. Such proteomics data can be acquired by sampling the experiment prior to phosphopeptide enrichment or be extracted from the non-modified peptides that are sequenced after phosphopeptide enrichment. This is particularly advantageous when comparing datasets with very different amounts of perturbation on the phosphoproteome, as most statistical tests will in this case require more stringent cutoffs for significance in the more perturbed datasets. In a mildly or unperturbed proteome, however, the quantitative data is more comparable across datasets, and will therefore result in more similar stringency when acquiring a cutoff.

Therefore you can perform a statistical test, such as Significance A, on your proteome data and use the identified cutoffs to define regulated events in your phosphoproteomics data. However, caution has to be applied when using proteomics data which is not directly extracted from the phosphoproteomics experiment itself. Due to less complex sample preparation protocols, data distribution can be narrower, leading to an overestimation of regulated phosphopeptides compared to the proteome.

6.4.2 Additional Considerations

Defining a cutoff will deem all entries above the cutoff to be regulated, while everything below is unregulated. However, biological systems are not binary and events falling right below the cutoff might still be regulated and biologically relevant. Therefore it is often advisable to have cutoff for the background as well as for the foreground. Events falling into the "grey zone" between those two cutoffs will be considered neither regulated nor unregulated and therefore not used for the analysis. Commonly only the ratios

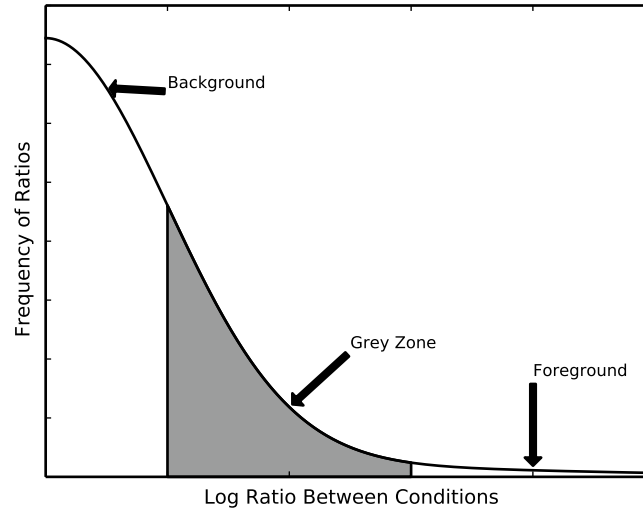


Fig. 2 Data separated into a Foreground and Background, the Grey Zone data is not used for further analysis, as it contains a mixture of regulated and unregulated peptides

within 1 standard deviation of the sample mean are used as the background cutoff, thus if a cutoff of 3 standard deviations has been chosen, then peptides with a standard deviation between 0 and 1 will become the background dataset and everything with a standard deviation above 3 will become the foreground dataset, as visualized in Fig. 2.

Considering the biological question of a specific analysis type, the user may benefit from performing all downstream analysis twice: once for the upregulated and once for the downregulated phosphosites. This is of particular relevance if a given dataset is skewed towards upregulation or downregulation, e.g., in the context of kinase or phosphatase inhibitors.

7 Platforms for Phosphoproteomics Analysis

Many software packages have been developed to search, filter and/or quantify (phospho)peptides. Below we present a few popular software packages that can perform all three steps of the analysis. All presented software is very mature, and the choice of software package is therefore usually based on user preferences. Generally the software can be classified into two different categories:

1. Pipeline/workflow oriented tools offer very high levels of flexibility and automation and are capable of creating workflows that can be reused for different types of follow-up analyses. However, generation of the workflow requires time and familiarity with the software, and can be difficult for first-time users.

2. Conventional software packages are generally equally powerful and tend to be easier to learn; however, they offer a lower degree of automation and flexibility. For those software options we put an emphasis on MaxQuant, which is freely available and very user friendly.

7.1 Workflow- and Pipeline-Based Software

Workflow-based software packages allow the users to generate a workflow, in a drag-and-drop manner. Once a workflow has been generated, it can be saved and shared among coworkers ensuring that everybody analyzes the data in the exact same way. Pipeline software packages work like workflows, but are generally made up of small command-line programs that can be chained together to form a pipeline by scripting.

7.1.1 Trans-Proteomic Pipeline

The Trans-Proteomic Pipeline [24] (TPP) is an open-source project from the Seattle Proteome Center (SPC). TPP is a web-based front end to a large collection of command-line tools. It can be used for almost any combination of MS instruments and labeling techniques.

A typical approach for using this pipeline in processing of quantitative phosphoproteomics data would include the following steps:

1. Standard input: Convert vendor format to mzXML (mzML or mzData)
2. Peptide assignment: Search data against one of the following databases: SEQUEST [17], MASCOT [18], COMET [25], ProbID [26] X!Tandem [13], or any other database of interest.
3. Validation:
   (a) Rank (phospho)peptides based on scores and filter based on a user-specified FDR using PeptideProphet [27].
   (b) Optionally: assemble peptides into proteins using ProteinProphet [27]
4. Quantification: use ASAPRatio [28] to calculate peptide ratios.

The software is very easy to install for Windows. While, Linux is officially supported it requires editing the *make* file to compile the source code. The advantage of this platform is that it covers all steps in phosphoproteomics data processing, from format conversion to quantification. This platform could be useful for inexperienced users in phosphoproteomics data analysis, as it offers pipelines that guide the user through the workflow.

7.1.2 TOPP and TOPPAS

The OpenMS [10] Proteomics Pipeline [11] (TOPP) is a large collection of programs with a command-line interface that can be chained together, much like the TPP. TOPPAS [9] is a workflow-

based graphical front end to TOPP. TOPP/TOPPAS can use the following search databases: Mascot [18], MyriMatch [15], OMSSA [14] and X!Tandem [13]. TOPP and TOPPAS are available for Windows, Linux, and OS X. For working with this platform you can follow the same four steps as described above for TPP. However, it allows for a greater level of flexibility in setting up the workflow.

7.1.3 Proteome Discoverer™

Proteome Discover is commercially available Windows software developed by Thermo Scientific. Its main focus is on data generated with Thermo Scientific Orbitrap instruments. Like TOPPAS it is also used in a workflow manner and can supports most of the popular search databases and labeling techniques.

7.2 Conventional Software Packages

Conventional software packages offer a workflow, which cannot be edited. However, these platforms generally allow a great level of flexibility within this scheme. This approach is more readily accessible as the platforms are very user-friendly and recommendable for beginners.

7.2.1 Mascot

Mascot [18] was developed by Matrix Science and is one of the oldest and most well-established database search engines. Mascot refers to the core database search algorithm, but complete data processing requires two main products: Mascot Server, which does the database search, and Mascot Distiller, which can do validation and quantitation. Mascot is only available for Windows.

7.2.2 MaxQuant

MaxQuant [23] is freeware developed at the Max Planck Institute of Biochemistry. MaxQuant uses a search database, Andromeda, developed specifically for this platform [16]. In recent years, the development of this platform has focused on including features that allow for compatibility with many different MS instruments (Thermo *.Raw, Brucker *.d, Sciex *.wiff and mzXML) and labeling techniques. This particular advantage renders MaxQuant an attractive software for the broad MS user community.

MaxQuant offers the option to group your input data, so that the chosen parameters can be applied specifically for a group or globally for all the data analyzed. As such, the user can analyze proteomics and phosphoproteomics data in parallel by applying specific parameters to some of the data (such as the search for phosphorylation in the phosphoproteome) while still using shared parameters, such as FDR cutoffs and the FASTA file being queried. Additionally, MaxQuant allows the option to configure the Andromeda database search engine, for example to include new modifications and FASTA files of interest. MaxQuant is only available for Windows.

8 Protocol: Phosphoproteomics Analysis with MaxQuant

8.1 Materials

To use MaxQuant, download the latest version from http://www.maxquant.org/. This protocol is designed for v. 1.5.0.0 but is readily transferrable to other versions.

Download a relevant proteome in FASTA format (*see* **Note 1**), for example from ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/proteomes.

8.2 Method

1. Launch MaxQuant. The user is presented with the interface seen in Fig. 3.
2. Navigate to "Raw files" tab.
3. Under *Input data* click "Load" (or Load folder) to import MS data files (or complete folders containing raw MS data) into MaxQuant.
4. Highlight all files that belong to the same experiment (usually also grouping replicates), and under *Edit exp. Design* click "Set Experiment." Write a descriptive name for the given experiment in the popup menu. All files should be assigned to an experiment.
5. If the files include different experimental workflows that require applying different parameters, the files must be grouped accordingly (*see* **Note 2**). Highlight the files to be grouped

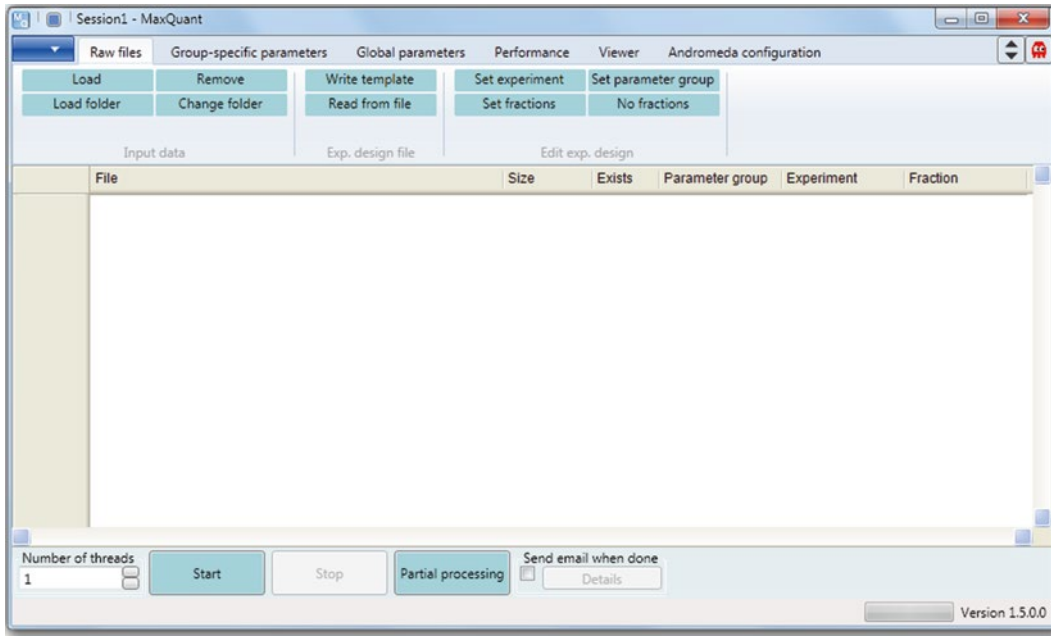


Fig. 3 Screen shot of MaxQuant version 1.5.0.0

together, by clicking "Set parameter group" and indicate a group number.

6. Navigate to *Group-specific parameters* tab.
7. For every parameter group repeat the following steps.
8. Under *Parameter section* select "General."
9. Under *Type* change the multiplicity (number of labels) to reflect the numbers of conditions in your experiment and select labels appropriately (*see* **Note 3**).
10. In the *Variable modifications* section keep default settings and scroll down to select modifications of interest if applicable. For phosphoproteomics experiments chose "Phospho (STY)" then click the ">" button.
11. In the *Digestion mode* section select the relevant enzyme using the ">" and "<" buttons (*see* **Note 4**).
12. Under *Parameter section:* select Instrument, and change values to reflect the quality and settings of the instrument that generated the MS data. Default parameters are provided for Orbitrap, Brucker TOF and AB Sciex TOF.
13. If applicable select "LFQ" in the drop-down menu, under Label-free quantification in the *Parameter section*.
14. Navigate to *Global Parameters* tab.
15. Under Parameter section select "General."
16. In the *FASTA files* section click "Add file" to important the proteome FASTA file.
17. Under *Identification* change PSM FDR (FDR at the spectrum level), Protein FDR and Site decoy fraction (modified peptides FDR) (*see* **Note 5**).
18. Navigate to *Performance* tab.
19. In the footer of the program change the number of threads (cores) to be used for processing. This can be up to the number of cores on the computer (*see* **Note 6**).
20. In the footer of the program click "Start."

9 Notes

1. Not all FASTA files are configured in Andromeda. Under the "Andromeda configuration" tab, this can be checked, and file of interest can be configured for use in MaxQuant.
2. Examples of situations where grouping files to apply group specific parameters would be useful:
   (a) Some of the files contain data with two labels while others have three.

(b) Different variable modifications are desired for different files, as some are to be used for phosphoproteome determination and others for proteome.

(c) The MS data files were generated using different MS-instruments or instrumental settings.

3. A triple SILAC experiments requires a multiplicity of 3. These should be specified: labels-1: nothing indicated, labels-2: Arg6 and Lys4, labels-3: Arg10 and Lys8.
4. Here, indicate the enzymes used for digestion in the experimental workflow. This is typically Lys-C and/or trypsin for most phosphoproteomics workflows.
5. We suggest setting all FDRs to 1 %.
6. If your system has less than 1GB ram per core, set the number of threads to the same number of GB ram available. Regardless of processing power, if all cores are used the processing capacity of the computer will be consumed.

Acknowledgments

This work was in part funded by the Novo Nordisk Foundation Center for Protein Research [NNF14CC0001]

References

1. Cohen P (2002) The origins of protein phosphorylation. Nat Cell Biol 4(5):E127–E130
2. Hughes C, Ma B, Lajoie GA (2010) De novo sequencing methods in proteomics. Methods Mol Biol 604:105–121
3. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 11(4), M111.010587
4. Lam H (2011) Building and searching tandem mass spectral libraries for peptide identification. Mol Cell Proteomics 10(12) R111.008565
5. Eng JK, Searle BC, Clauser KR, Tabb DL (2011) A face in the crowd: recognizing peptides through database search. Mol Cell Proteomics 10(11) R111.009522
6. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1(5):376–386
7. Ong S-E, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 1(6):2650–2660
8. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak M-Y, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, 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 EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30(10):918–920
9. Junker J, Bielow C, Bertsch A, Sturm M, Reinert K, Kohlbacher O (2012) TOPPAS: A Graphical Work flow Editor for the Analysis of High-Throughput Proteomics Data. J Proteome Res 11(7):3914–3920
10. Sturm M, Bertsch A, Gröpl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-

Trieglaff O, Zerck A, Reinert K, Kohlbacher O (2008) OpenMS—an open-source software framework for mass spectrometry. BMC Bioinformatics 9:163

11. Kohlbacher O, Reinert K, Gröpl C, Lange E, Pfeifer N, Schulz-Trieglaff O, Sturm M (2007) TOPP–the OpenMS proteomics pipeline. Bioinformatics 23(2):e191–e197
12. Deutsch EW (2012) File formats commonly used in mass spectrometry proteomics. Mol Cell Proteomics 11(12):1612–1621
13. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20(9):1466–1467
14. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. J Proteome Res 3(5):958–964
15. Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis research articles. J Proteome Res 6(2):654–661
16. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805
17. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5(11):976–989
18. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551–3567
19. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4(3):207–214
20. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635–648
21. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature 473(7347):337–342
22. Wiese S, Reidegeld KA, Meyer HE, Warscheid B (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 7(3):340–350
23. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
24. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, Eng JK, Martin DB, Nesvizhskii AI, Aebersold R (2010) A guided tour of the trans-proteomic pipeline. Proteomics 10(6):1150–1159
25. Eng JK, Jahan TA, Hoopmann MR (2013) Comet: an open-source MS/MS sequence database search tool. Proteomics 13(1):22–24
26. Zhang N, Aebersold R, Schwikowski B (2002) ProbID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. Proteomics 2(10):1406–1412
27. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74(20):5383–5392
28. Li X-J, Zhang H, Ranish JA, Aebersold R (2003) Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. Anal Chem 75(23):6648–6657

Chapter 23

Systems Analysis for Interpretation of Phosphoproteomics Data

Stephanie Munk, Jan C. Refsgaard, and Jesper V. Olsen

Abstract

Global phosphoproteomics investigations yield overwhelming datasets with up to tens of thousands of quantified phosphosites. The main challenge after acquiring such large-scale data is to extract the biological meaning and relate this to the experimental question at hand. Systems level analysis provides the best means for extracting functional insights from such types of datasets, and this has primed a rapid development of bioinformatics tools and resources over the last decade. Many of these tools are specialized databases that can be mined for annotation and pathway enrichment, whereas others provide a platform to generate functional protein networks and explore the relations between proteins of interest. The use of these tools requires careful consideration with regard to the input data, and the interpretation demands a critical approach. This chapter provides a summary of the most appropriate tools for systems analysis of phosphoproteomics datasets, and the considerations that are critical for acquiring meaningful output.

Key words Phosphoproteomics, Systems analysis, Functional networks, Gene ontology, Sequence motifs, STRING, Cytoscape

1 Introduction

Mass spectrometry-based phosphoproteomics has proven a robust and reliable approach for analysis of site-specific protein phosphorylation in a multitude of biological settings. Global phosphoproteomics investigations can generate datasets consisting of tens of thousands of identified and quantified phosphorylation sites (phosphosites). Ultimately, the aim of performing such high-throughput phosphoproteomics investigations is to extract meaningful biological information that can provide mechanistic insights or fuel hypotheses for further studies. Although this remains the bottleneck in the field of phosphoproteomics, the continuous developments and improvements of bioinformatic tools within systems biology provide useful platforms to analyze and dissect large datasets. Such tools can be used to determine, which kinases are more active in a given context, which pathways or biological processes

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4_23,

are significantly enriched in the data, or to generate and visualize the data in the context of biological networks. In this respect, extensive lists and reviews of bioinformatics tools for analysis of large-scale datasets have been published [1–3]. When performing a thorough and appropriate analysis of the data it is very important to choose a means to visualize the results. This will provide the best foundation not only for interpretation but also for communicating key findings to the scientific community.

Phosphoproteomics data contain three levels of information that are useful for systems analysis, namely the individual phosphosites, the proteins on which these reside (phosphoproteins), and the individual phosphosite ratio (fold changes) between experimental conditions when performed quantitatively. Both the phosphoprotein and phosphosite information are relevant as input for different systems analyses, and for some tools both data types are valid. It is therefore important to determine the type of input data that is compatible with the given tool prior to performing an analysis and consider whether an analysis at phosphoprotein or phosphosite level would be most relevant for the overall investigation. Similarly, when both phosphosite and phosphoprotein data are applicable, a clear definition of the purpose of the analysis should determine the appropriate input to produce a relevant output.

Phosphorylation events dictate many protein functionalities, including enzymatic activity, interactions with other macromolecules, and subcellular localization. While systems biological analysis are very useful for providing a global understanding of large phosphoproteomics datasets, it is important to keep in mind that the function of the individual phosphosites cannot be determined with this type of analysis. Actually, the roles of most known phosphosites are not known. However, the information acquired from systems analysis can form the basis of hypothesis formulation and generation of concrete strategies for future experimental directions. This chapter provides an overview of the most commonly used tools for systems analysis of phosphoproteomics data.

2 Pathway and GeneOntology Analysis

Global phosphoproteomics screens are often performed to address the question of which biological processes or pathways are perturbed in a given set of conditions. The mere size of modern phosphoproteomics datasets renders this a daunting task if performed manually, making bioinformatics enrichment tools indispensable for data interpretation.

Enrichment analysis tools annotate proteins with different classes or categories from a list of interest according to a database(s) of choice. Subsequent statistical analyses then provide a quantitative measure of the enrichment (over- or underrepresentation)

of the pathways or processes in the queried data compared to a reference or background. In addition to fold enrichment, these tools provide a *p*-value which is commonly calculated using the Fisher's exact test. Correction for multiple hypothesis testing is usually performed using Benjamini–Hochberg or Bonferroni correction [3]. Importantly, some tools for pathway and ontology analysis allow the option to provide a reference dataset. For most biological studies it is recommendable in such analyses to compare the phosphoproteins of interest against a reference set of background proteins from the same experimental data. This reference dataset usually consists of phosphoproteins with unregulated phosphosites, although in some cases it can be beneficial to use the whole dataset. Applying a custom reference, rather than comparing experimental data to databases comprising complete genomes, compensates for biases that are invariably present in experimental data. In phosphoproteomics studies such bias arises particularly from protein abundances and from the method of sample preparation, which may enrich some fractions of proteins or peptides with specific chemical properties more efficiently than others. When applicable, a custom reference dataset therefore allows for more relevant interpretation of an enrichment analysis.

Gene ontology (GO) annotation and pathway enrichment analysis are typically conducted on the phosphoprotein level. However, phosphoproteomics studies offer information on the phosphosite level that can provide more contextually relevant results. The difference lies in that the use of phosphosites takes into account that the number of phosphoproteins does not directly translate to the number of phosphosites. As such, the proteins in the queried regulated data may be more heavily phosphorylated compared to those in the reference. This information is lost when using phosphoprotein data only. Thus phosphosite level data should be used if applicable. Another challenge in performing gene ontology analyses is the overwhelming numbers of functional annotations and great level of overlap between terms, such as "apoptosis," "regulation of apoptosis," and "regulated cell death", which poses a hurdle for interpretation for which there is still not a robust solution. However, the DAVID annotation tool (see below) provides a recommendable clustering option that can organize a heterogeneous set of annotation types, and ease the interpretation process.

When interpreting enrichment data it is important to realize that enriched pathways and GO categories are identified by statistical analysis that test for significance. Yet to extract functional meaning the user must interpret these results with a biologically critical approach. A pathway with three proteins identified in the queried dataset may be statistically significantly enriched, but the user must question whether three proteins are enough to form the foundation for any biological conclusions regarding perturbed pathways. Such cases may still be useful if the user finds biological relevance in

the pathway or the proteins are known to be involved in the given biological context. However, for global interpretations it is recommendable to require a minimum number of identified members for each category to be included in further analysis. From the results, the user will often identify enriched categories that are relevant and maybe even expected for the biological experiments performed. These categories are still useful as they provide the user with a list of proteins from the dataset that are involved in these processes of interest. However, it is the unexpected findings that generate new models and concepts, and must therefore also be considered carefully, and subsequently subjected to thorough investigation.

It is also important to bear in mind the mere nature of annotation databases when interpreting ontology and pathway enrichment analysis. Most databases are manually curated and the ongoing annotation of proteins in these databases relies on experimental findings from scientific literature [4]. The types of experiments conducted and their biases will therefore propagate into databases using these to assign annotations. Categories that are more extensively studied will inherently be more thoroughly annotated, and are therefore more easily mapped in the analysis of high-throughput datasets. Enrichment analysis of phosphoproteomics screens is therefore susceptible to this bias, stressing the importance of manual inspection of the results by the user.

The output for most ontology and pathway enrichment tools is in the form of a list. This is a rich source of information, and the user may be interested in just one category or the most enriched categories. Yet these lists can be overwhelming to look through, and for global interpretation it is beneficial to represent the data in an appropriate visual manner. Most commonly, pathway and ontology enrichment are visualized as bar graphs where the bar height represents fold enrichment or log-transformed *p*-values. Alternatively pie charts are also used, showing the percentage of genes mapping to the annotations. For comparison of multiple dataset, for example different stimuli or dynamic studies with several time points, clustering of the annotations based on their *p-values* can provide a useful means to visualize and extract clusters of annotations that are unique or common to the variables compared. A few of the more commonly used databases for gene ontology and pathway analysis are briefly described here, as well as an example of a tool to perform these. Other tools are listed in Table 1.

2.1 Databases

The *Gene ontology* [5] project (http://www.geneontology.org) is a great biological effort that describes the biology of a gene product in three structured vocabularies (ontologies): biological processes, cellular components, and molecular functions. GO terms exist on different levels, from the very general terms (first level of GO) to very specific terms (*see* Fig. 1). This database is queried by most enrichment tools.

Table 1
Tools for GO and pathway analysis

Name	Description	Cost	Link
DAVID [8]	Annotation and enrichment of GO, Pathways, and many more	Free	http://david.abcc.ncifcrf.gov/
InnateDB [28]	Annotation and enrichment of GO, Pathways. Visualization	Free	http://www.innatedb.com/
KEGG Atlas [29]	Visualization in KEGG pathway maps	Free	http://www.kegg.jp/kegg/atlas/
GoMiner [30]	Gene ontology analysis	Free	http://discover.nci.nih.gov/gominer/
MetaCore	Functional analysis and visualization of omics data	$	http://thomsonreuters.com/metacore/
Ingenuity pathway	Annotation and visualization of networks and pathways	$	http://www.ingenuity.com/

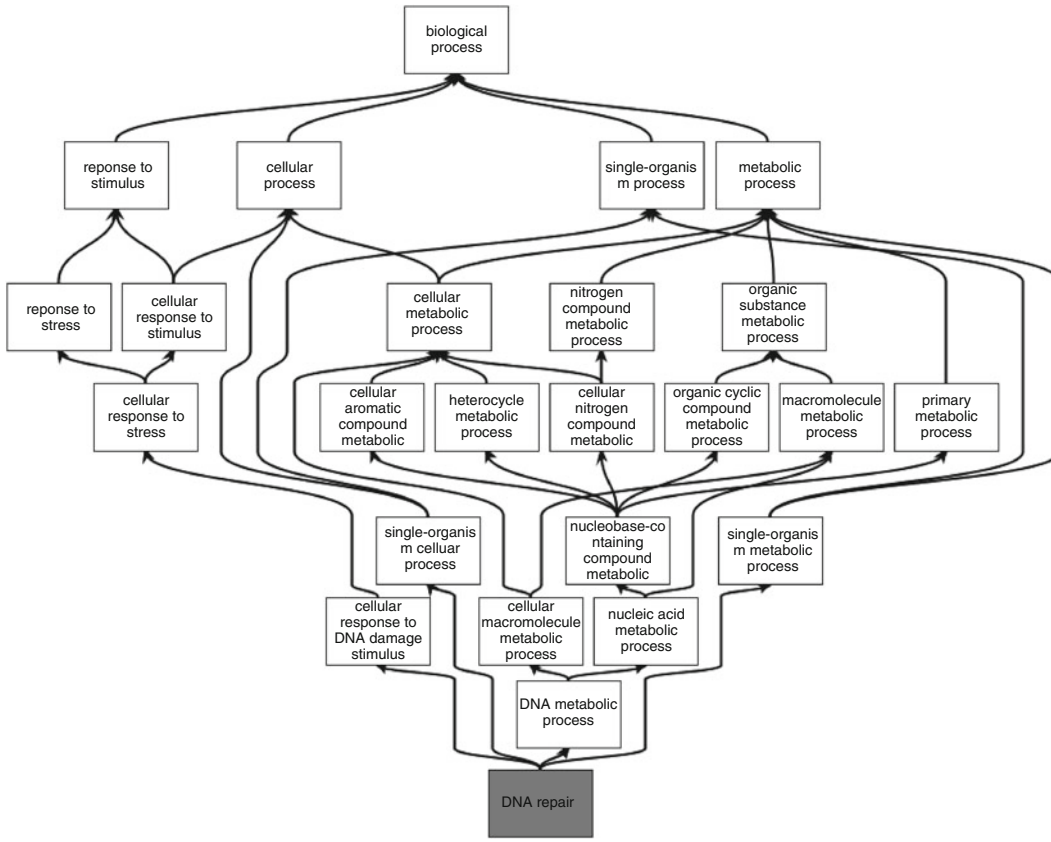


Fig. 1 GO term hierarchy. The hierarchy of the GO biological process "DNA repair." This and other GO term hierarchical relations can be retrieved here: http: www.ebi.ac.uk/QuickGO/

KEGG [6] (Kyoto Encyclopedia of Genes and Genomes) is a collection of manually curated databases that integrate genomic, chemical, and systemic functional information. The KEGG PATHWAY database is a collection of manually drawn maps integrating information about genes, proteins, chemical compounds, and many more. KEGG covers pathways within metabolism, cellular processes, organismal systems, human disease, drug development, genetic information processing, organismal systems and environmental information processing (http://www.genome.jp/kegg/).

Reactome [7] is a curated pathway database, which aims to provide a bioinformatics tool for visualization, interpretation, and analysis of pathways. The database is based on manually annotated pathways from textbooks and scientific articles, to map "reactions" that include binding, activation, translocation, degradation, and biochemical reactions (http://www.reactome.org/).

2.2 DAVID

DAVID (Database for Annotation, Visualization, and Integrated Discovery) [8] is a web-based bioinformatics resource (http://david.abcc.ncifcrf.gov/home.jsp) with a comprehensive repertoire of functional annotation options, that renders it recommendable for large phosphoproteomics data analysis. DAVID queries databases for enrichment of GO terms, pathways (including KEGG, Reactome, and BioCarta: http://www.biocarta.com), disease (such as OMIM: http://omim.org/), protein domains (PFAM [9], SMART [10], and many more) and other categories that may be relevant for phosphoproteomics screens. All annotations are assigned to DAVID identifiers that nonredundantly represent each gene, gathering gene identifiers and accessory information from various databases. DAVID has four main analysis features: "Functional annotation," "Gene functional classification," "Gene ID conversion," and "Gene name batch viewer." "Gene name batch viewer" allows users to browse their gene list of interest, where each gene can be clicked on to acquire more information of gene products, related references and the option to identify related genes. "Gene functional classification" groups functionally related genes into groups, and provides an enrichment score along with links to view their biological relations which can be visualized as heat maps. The "functional annotation" feature encompasses the analysis for annotation enrichment that is truly valuable for interpretation of large-scale phosphoproteomics data. The *p*-values for the annotated categories are calculated by EASE score, which is a modified Fisher's exact *p*-value.

To perform functional annotation analysis, the user must provide a gene list of interest. This is done in a user-friendly interface, allowing the user to either paste in the gene list or upload a file that may contain many lists of interest (multi-list file). In either case the format must be one gene per row, and the entries must be specified as either a "Gene" list or "Background." From the "List manager" in the "List" tab the user can specify which of the input gene lists

(if more than one has been entered) should be queried. In the "List manager" it is also worth noting how many of the uploaded genes make up the lists, as the mapping of entry genes is rarely complete. In the "Background" tab, the user can choose whether to use an uploaded custom list as background, or the complete genome from a specified species (genes from microarray chips are also provided as a background option). Next, the user is offered various analysis and visualization options in the "Annotation summary results." Most importantly, the user must carefully choose the annotation types of interest from a comprehensive list. Under each annotation category the different database options are also accompanied by the percentage of the genes in the list that annotate to that particular database, the number of genes that this encompasses, and the option to view a chart report for the specific database. When the databases of interest have been chosen, the user can view the combined results as Functional annotation clustering, Functional annotation chart, or Functional annotation table. *Functional annotation chart* provides a list of the enriched functional annotations, supplemented with *p*-values, percentage (and number) of the uploaded genes mapping to the annotation, and hyperlinks to access the list of genes classified in each annotation, or to browse related terms. *Functional annotation clustering* groups annotations based on similarities of the genes classifying to the categories, with the aim to reduce redundancy. This tool also provides *p*-values, percentage (and number) of genes mapping to the annotation, and hyperlinks to access the list of genes classified in the each annotation, or to browse related terms. The functional annotation chart and clustering tools are equipped with the option to choose the level of stringency, *p*-value cutoff, and count threshold. These options can be very useful when the list of enriched terms is overwhelmingly long, and intuitively, more stringent analysis cutoff thresholds should give rise to more confident analysis results. *Functional annotation table* queries the DAVID knowledge database, and provides the annotations for each uploaded gene, in a table. All types of analysis performed in DAVID can be saved to a local drive.

As mentioned above, enrichment analysis of phosphoproteomics data can benefit from including phosphosite information. To calculate enrichment on the phosphosite rather than the phosphoproteome level, the user must perform the enrichment analysis for the genes of interest in DAVID using the custom background list of genes as reference, as described above. An additional dataset with annotations of all the genes in the reference dataset must be acquired using the Functional annotation table option. The sum of the phosphosites mapping to all the proteins of each enriched category and for all proteins annotated within those categories in the reference dataset can then be used to calculate fold enrichment and *p*-values with the Fisher's exact test. This is a tedious task if performed manually, and we have therefore generated a tool to allow for this analysis.

2.2.1 Protocol of Enrichment Analysis on the Phosphosite Level

The purpose of this protocol is to allow the user to perform enrichment analysis based on the number of phosphosites in regulated and reference datasets, rather than the number of phosphorproteins. For this, the phosphoproteins are used for annotation of categories of interest, based on the DAVID tool. At this time, our tool supports the following DAVID categories: OMIM, GOTERM_BP_FAT, GOTERM_CC_FAT, GOTERM_MF_FAT, BIOCARTA, KEGG_PATHWAY, and REACTOME. In due time, as the tool is updated other categories will be included, and all that are supported will be listed at phomics.jensenlab.org/phospho_enrichment where the tool is available.

2.2.2 Material

This protocol requires two datasets, one for the regulated phosphosites and one for the unchanged phosphosites. These datasets should include an identifier supported by DAVID for each phosphoprotein, and the phosphosites. The file should be in a tab separated in the following format:

[Uniprot ID] [Tab] [Site], for example: P23443 T233.

2.2.3 Method

1. Open the DAVID website: http://david.abcc.ncifcrf.gov/home.jsp.
2. Click "Functional Annotation."
3. Copy in the list of identifiers for the regulated dataset in the text box designated "Step 1".
4. Choose the identifier used under "Step 2".
5. Indicate under "Step 3" that your dataset is a gene list.
6. Click "Submit list." This will bring you to the "List" tab where your list of entries will be designated a default name such as "List 1".
7. Choose the categories of interest under each main header.
8. Click on "Functional Annotation table".
9. A new window will pop up. Right click on the disk symbol next to "Download File".
10. Chose "Save link as". Name it FwDAVID
11. Repeat **steps 1–10** for the reference list. Name results file BgDAVID
12. At this point the user should have 4 files: the two original datasets for regulated and unchanged phosphosites as indicated under "Material" and two results files from DAVID.
13. Go to phomics.jensenlab.org/phospho_enrichment
14. Upload the four datasets where indicated.
15. Select the method for multiple testing. We recommend FDR.

16. Change alpha if necessary, (acceptance threshold for correcting for multiple testing). This is the same as the "corrected *p*-value" cutoff.
17. Click "Submit."
18. The result page will have the following columns:
    - ID: The category term
    - Description: Description of the category
    - p value: uncorrected *p*-value
    - p Bonferroni: Bonferroni corrected *p*-value
    - p Holm: Holm–Bonferroni corrected *p*-value
    - p Sidak: Sidak corrected *p*-value
    - FDR: FDR corrected *p*-value

2.3 Interpretation

The interpretation of the results from this analysis is to be performed in the same manner as a classical enrichment analysis performed at the protein level as described above. For large datasets, the output comprises a very long list. As previously described, this list must be critically and stringently filtered using for example *p*-value cutoffs and requiring a minimum number of annotated members.

3 Visualizing and Analyzing Biological Networks

Functional and protein–protein interaction networks are becoming an integrated part of high-throughput data analysis as they provide a systems-level context of the individual components of large datasets. For phosphoproteomics data, networks are generated from a set of phosphoproteins and will therefore bear no information on specific phosphosites. The proteins are visualized in the network as nodes that are connected to each other by edges (*see* Fig. 2). Edges can represent physical interactions, enzymatic reactions or a functional connection, the evidence for which arises from various repositories of experimental results and from text mining of the literature. The type(s) of edges represented in a network is determined by the tool and database utilized and most tools offer the option to designate which types should be included in the analysis.

While a network visualizes how the given set of proteins are interconnected, drawing meaningful contextual conclusions or forming hypothesis for further studies, ultimately relies primarily on the user's biological expertise. Yet, certain network topology features can be useful to guide the user in this process. *Hubs* are highly connected nodes that are vital for the stability of the network and are often essential and conserved genes in model organisms [11, 12]. Hubs can be considered functional centers

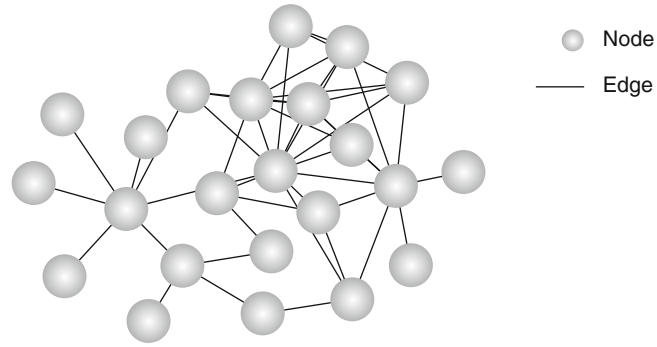


Fig. 2 Network. Example of a simple biological network. The *circles* are called nodes and represent each gene/protein used to generate the network. The lines connecting the nodes are termed edges and represent the association between the nodes

either inside modules to coordinate a biological process, or to link different modules [11]. Furthermore, some nodes will show a high level of interconnectivity and form what is known as *clusters*. The nodes within a cluster are often involved in the same process, which the user must discern, and tools are available to extract these clusters. More advanced analyses, such as visualization and enrichment of GO terms are also very informative when performed in the network context, which is a feature not available in most other enrichment tools.

To optimize the gain from a network analysis, there are certain considerations the user can take concerning the input data, always keeping in mind the relevance for the overall biological investigation. As such, a network generated from thousands of phosphoproteins will typically yield a very large and complex network graph which will require extensive processing to decipher. In such circumstances, the user may benefit from using a subset of the data, such as phosphoproteins with regulated phosphorylation events (*see* Chapter 22) or those belonging to a significantly enriched GO category. On the contrary, if the phosphoproteomics data yielded a very restricted number of phosphoproteins, the user may be interested to explore how these few nodes are involved in a larger context. These circumstances could benefit from importing neighboring nodes from databases to the network, a feature that is available in most tools.

Here we will describe the most commonly used, free, web-based and downloadable software tools for network generation and analysis. Others tools are listed in Table 2.

3.1 STRING

STRING (Search Tool for the Retrieval of Interacting Genes) (http://string-db.org/) [13, 14] is a web-based database resource that maps known and predicted interaction, either direct (physical) or indirect (functional), for a query gene or list of genes. Currently, this meta-resource provides a comprehensive coverage of 1133 organisms [14].

Table 2
Tools for network analysis

Name	Description	Format[a]	Link
STRING [13]	Database repository for network mapping	Web	http://string-db.org/
Cytoscape [16]	Network analysis. Expanded by apps	Local	http://www.cytoscape.org/
VisANT [31]	Network and pathway construction	Web	http://visant.bu.edu/
Biological Networks [32]	Integrative network analysis	Local	http://biologicalnetworks.net/
VANTED [33]	Visualize biological networks and hierarchies	Local	http://vanted.ipk-gatersleben.de/

[a]Web: Web-based tool. Local: downloaded to local drive

The STRING database generates networks based on functional associations [15]. This means that the edges connect proteins which contribute to the same functional process. The protein–protein associations are based on genomic context (gene proximity, gene fusion, co-occurrence, and co-expression), high-throughput experiments, and preexisting knowledge (curated databases and text-mining). Advantages of the STRING database resource are that it is very comprehensive, provides edge confidence scoring and has an interactive interface [13]. The web page also offers a useful help center, which provides beginners with a solid foundation to get started.

3.1.1 Input Data

As previously mentioned, network analyses for phosphoproteomics data is performed on the protein level. STRING supports a number of identifiers including gene names, protein names, and various accession codes. In addition to querying a list of proteins, the user can enter one protein of interest, in which case STRING will provide a network with predicted interactions partners (*see* Fig. 3). In case an input identifier can be mapped to multiple proteins, the user has the option to choose the protein of interest.

3.1.2 Network and Protein Associations

When a list of genes is provided, STRING generates a network based on interactions between proteins in the input data only, with the option to include predicted interactors during analysis. When only one protein is provided, STRING generates a network based on predicted interactors that can be expanded or decreased. The resulting network is displayed in a user-friendly interface, with many features that aid the interpretation process. Clicking on the nodes prompts a pop-up box with a series of accessory information that can help the user to navigate the proteins in the network. Edges also comprise a rich source of information. All the evidence supporting the association can be viewed in the "Evidence" mode,

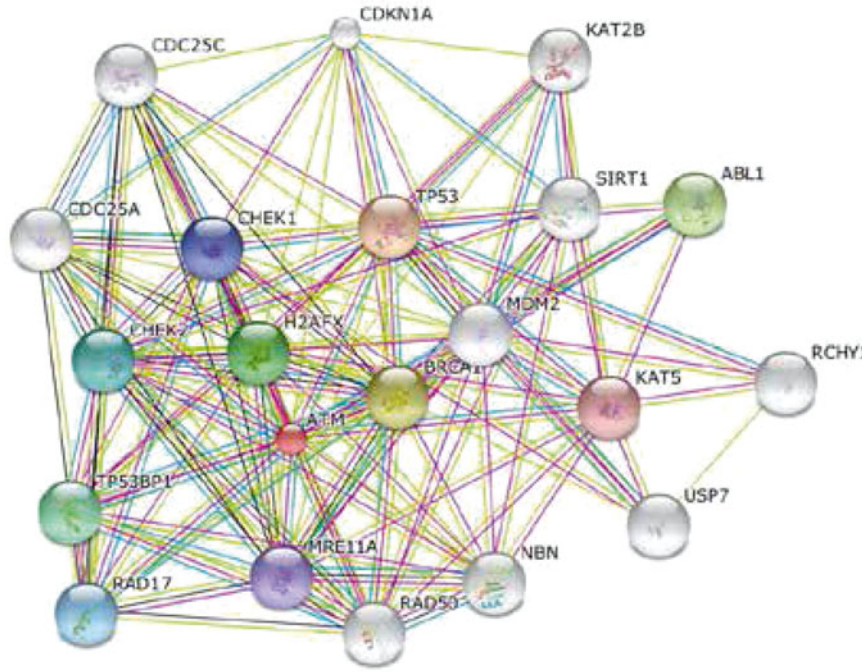


Fig. 3 STRING. Example of a network generated in STRING. The different *line colors* represent the different types of evidence associating the nodes

and extensive descriptions are available by clicking on the edges. Each edge is assigned a confidence score representing the probability of the association between two nodes, which is calculated from the scores of all the evidence types therein included. This can be visualized in the "confidence" view, where thicker edges denote more confident associations. The user is also provided with the option to personalize the criteria for the edges, such that the evidence types of interest and the level of confidence of the edges required for inclusion in the construction of the network can be customized. This is particularly useful for reducing complexity of large networks, and is an advantage of STRING compared to other network analysis tools. Networks can also be expanded by including predicted interactions, offering a framework to explore the context of proteins beyond the queried dataset.

The multitudes of options to process the network in STRING provide the foundation for the level of interpretation that the user desires, which must be complemented by the stringency of the criteria implemented. More lenient criteria are beneficial for the user seeking broader overview of the potential associations whereas users interested in more confident interpretations for concrete targeted follow-up experiments are better served with a more stringent approach.

3.1.3 Network Analysis Options

In "Advanced" mode several tools are available to perform more elaborate analysis. These are particularly applicable for interpretation of discovery phosphoproteomics data. The enrichment analysis tool provides a list of overrepresented biological functions and pathways (Gene Ontology, KEGG, Pfam, and InterPro) that can be browsed and viewed directly in the network. A clustering analysis option groups highly connected nodes and repositions them accordingly. While the biological meaning of clusters in other network analysis tools relies on the user's biological expertise for interpretation, by using the enrichment and clustering options together in STRING, the user can examine the clusters for enriched categories.

3.1.4 Export Options

STRING is also comprehensive with regard to export options including high and low resolution graphics, text files with network details, and xml summary files.

3.2 Cytoscape

Cytoscape [16] (http://www.cytoscape.org/) is an open-source platform for visualization, integration, and analysis of biological networks, and is compatible with Windows, Mac, and Linux operating systems. The advantage of Cytoscape over other network analysis tools lies in the ability to integrate many levels of information (attributes), either provided by the user or imported from databases. Particularly the option to import node and edge attributes from different sources renders this tool particularly advantageous for phosphoproteomics data interpretation.

The power of Cytoscape lies in the accessible application programming interface (API) that allows users to write application extensions to the software, called Apps (known as plugins prior to v3.X). Apps extend the functionality of Cytoscape, and allow the user to perform analyses of network features, such as clustering, and query the network for biological functionalities such as GO terms. Apps can be accessed by other users by direct download from the Cytoscape interface or through the Cytoscape webpage. Cytoscape recently released a new and improved version of the software, v3.X. Notably, this version does not support apps from the previous versions (v2.X). While the number of apps for v3.1 is rapidly growing, the repository of v2.X compatible apps is to date more extensive.

3.2.1 Current Cytoscape Versions

V3.X is the latest version of Cytoscape, offering new features including a new user interface. As mentioned before, this version does not support Apps from the previous v2.X, and therefore the user must consider the version of Cytoscape to use before commencing analysis. Fortunately Cytoscape sessions saved in v2.X can be opened in v3.X. But it is important to note that the opposite does not apply, and if a v2.X session file is opened in v3.X, it is no longer compatible for use in v2.X. Cytoscape v2.X is being maintained, and eventually v3.X will be comprehensive.

3.2.2 Creating a Network

Cytoscape supports many standard network and annotation formats, including SIF (Simple interaction Format), PSI (Proteomics Standard Initiative), tab delimited text files, and MS Excel™ Workbook.

There are four ways of generating a network in Cytoscape: (1) importing preexisting, formatted network files, (2) importing preexisting text or Excel files, (3) importing data from a database, or (4) manually creating a network by adding nodes and edges. Network file formats supported by Cytoscape include, SIF (Simple Interaction Format), GML (Graph Markup Language), PSI (Proteomics Standard Initiative), delimited text files, Excel workbooks, and many more. Importing networks from unformatted text and Excel files simply requires the user to specify the columns that contain the source and target node and the interaction information. This option allows STRING generated networks to be imported into Cytoscape using several formats including PSI and txt (text summary file). V3.X also provides the option for the user to query a list of proteins in a comprehensive repertoire of public databases, including MINT (Molecular Interaction) [17], STRING, and Reactome to retrieve protein–protein associations. For v2.X a simplified version of this function is available, and interaction mapping for lists of proteins is performed using specialized plugins such as MiMI [18] and StringWSClient .

3.2.3 Customizing Networks

As is the case for most bioinformatics tools, Cytoscape was not developed for interpretation of phosphoproteomics data, and most available Apps only provide analysis and information on the protein level. To visualize phosphosite specific information, such as fold change, the user must import these as node attributes that can subsequently be visualized in the network (*see* Fig. 4). In this context, the option to manipulate various features of the network style in Cytoscape is particularly useful for data interpretation. Users can choose to map any appropriate attribute to nodes or edges, as desired. One useful strategy for phosphoproteomics is to designate unique colors for proteins with regulated and non-regulated phosphosites (*see* Fig. 4). Color gradients representing the least to the most regulated phosphosite ratios can also help visualize patterns in the regulation of the phosphosites (*see* Fig. 4). Different types of protein–protein associations can also be visualized by different types of lines representing the edges. For example, solid lines could represent physical interactions, and dashed lines enzymatic associations.

3.2.4 Analyzing Networks: Apps (Plugins)

Interpretation of the visually customized networks relies largely on manual inspection by the user, which is particularly tedious for large networks. Combining visualization of variables from the phosphoproteomics data (such as fold change) with network or biological information can be very advantageous to extract

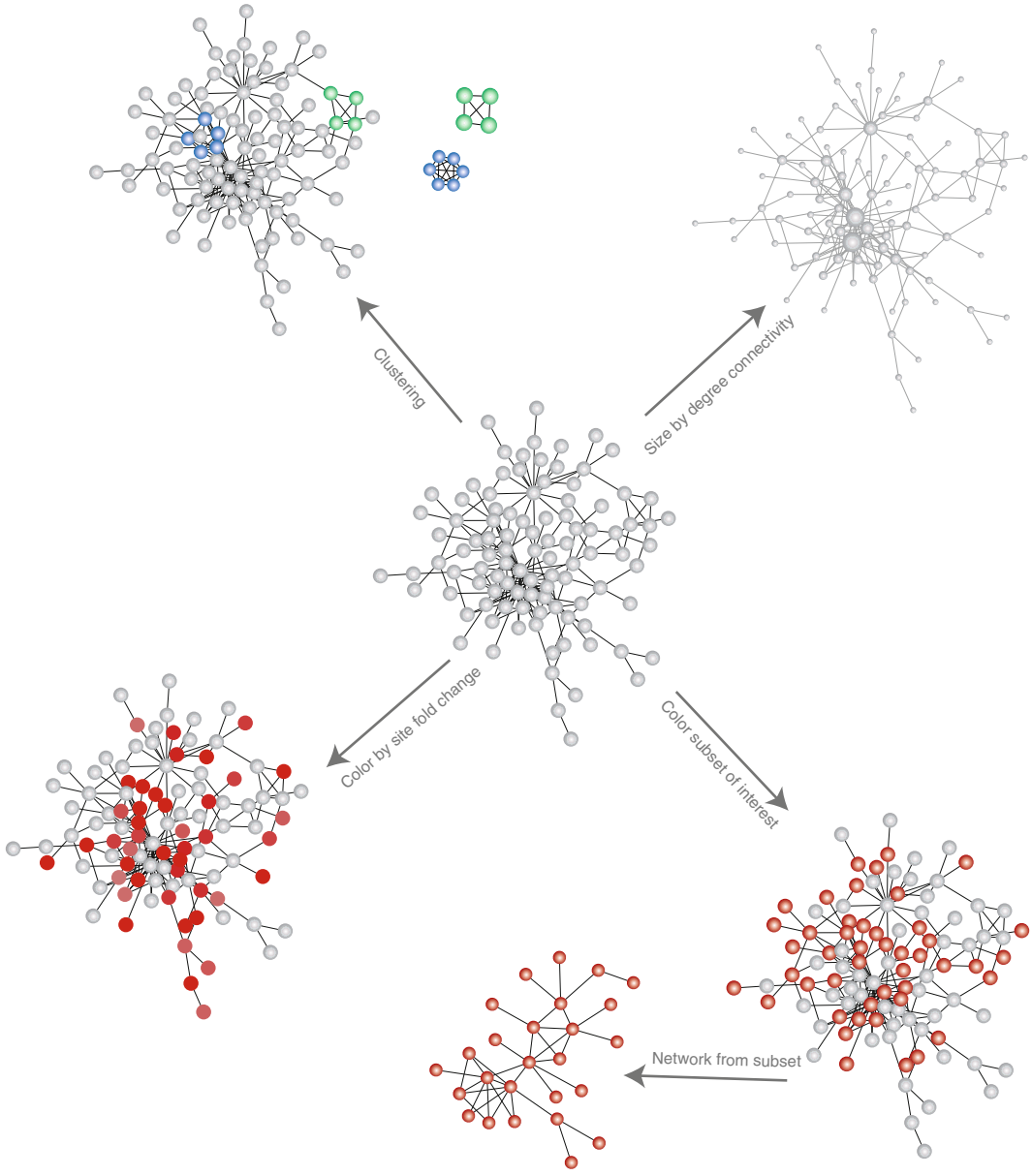


Fig. 4 Cytoscape. Interpretation and analysis of network are best served by manipulating network features, reducing network complexity, or visualizing attributes of interest

biological meaning. For this, a multitude of Apps are available that generally allow for analysis of the network topology or for associating biological functionalities of interest. Cytoscape apps are mostly well documented and reviewed, and a few are briefly summarized here (Table 3).

An integrated *NetworkAnalyzer* tool performs a comprehensive statistical analysis of the network topology, such as degree distribution and neighborhood connectivity, which can be visualized

Table 3
Cytoscape apps

Name	Description	Link	Version
MCODE [19]	Clustering	http://baderlab.org/Software/MCODE	2.6–3.X
ClusterOne	Clustering	http://www.cs.rhul.ac.uk/home/tamas/assets/files/cl1/cl1-cytoscape-0.1.html	2.6–3.X
jActiveModules	Clustering	http://wiki.cytoscape.org/Presentations/07_Complexes	2.5–3.X
ClueGO [22]	GO and pathway analysis and enrichment	http://www.ici.upmc.fr/cluego/	2.8–3.X
BinGO [23]	GO annotation and enrichment	http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html	2.6–3.X
ReactomeFIPlugin [7]	GO annotation and pathway analysis	http://www.reactome.org/	3.X
KEGGscape	Map data to KEGG pathway	https://github.com/idekerlab/KEGGscape	3.1
CyTargetLinker [34]	Map regulatory interactions	http://projects.bigcat.unimaas.nl/cytargetlinker/	2.8–3.X
WordCloud	Summarize network function as logos	http://baderlab.org/Software/WordCloudPlugin	2.6–3.1

in the network. These automatically become node and edge attributes. Particularly useful for biological networks is the "degree distribution" (*see* Fig. 4), a parameter that describes the number of edges connected to a node, and is easily visualized as node size. Combining this with color scheme designating phosphoproteins with regulated phosphosites for example, allows the user to determine whether certain regulated proteins are hubs.

1. *MCODE* [19] clusters networks based on topology to find groups of highly interconnected nodes, which will often partake in a common complex or pathway (*see* Fig. 4). For networks containing expression data, *jActiveModules* based on a model by Ideker et al. [20], can be used to identify sub-networks with significantly high or low average expression (*p*-values for the expression data must therefore be imported as attributes). This app is useful for phosphoproteomics data as it allows the application of quantitative phosphosite information with network topology to find clusters with similar phosphosite regulation patterns.
2. *ClueGO* [21] generates a functional network in which genes are clustered according to GO, KEGG and BioCarta categories of the user's choice. ClueGO can be applied to a set of proteins or to a preexisting network, and the resulting GO term clusters are grouped according to similarities in the categories of the associated genes. ClueGO also calculates term and group significance (*p*-values and corrected *p*-values). This app can be used in combination with the *GOlorize* app to map the biological functions back to a Cytoscape network with the genes of interest. *CluePedia* [22] extends the ClueGO functionality by allowing integration of other biological data, such as expression data, that can be used for further statistical analysis. CluePedia also provides an intracellular pathway-like Cerebral layout.
3. *BinGo* [23] is another popular app for GO annotation and enrichment analysis, which can also be performed directly from a network or with a list of input genes. This tool displays not only the overrepresented (or underrepresented) categories but also their parent categories.

Large quantitative phosphoproteomics investigations often invite for speculations about whether the phosphoproteins containing regulated phosphosites or the overrepresented pathways have some relevance in disease. The *ReactomeFIPlugin* [7] produces a functional network that can be queried in the Reactome database to identify sub-networks and annotate these for enriched pathways or disease genes [24, 25]. A Reactome pathway of choice can also be opened as a diagram and the nodes present in the Cytoscape network are highlighted. The user can also view all nodes annotated to a disease of interest directly in their network.

3.2.5 Saving and Exporting

Networks generated in Cytoscape can be exported as graphics in formats such as JPEG, PDF, PNG, Post Script, or SVG format. The Cytoscape session can also be saved for continued analysis at a later time point. It is important to note, that not all the information generated with different apps are stored in these session files. Also, the data from the session can be exported as a table.

Interpretation of phosphoproteomics data can be greatly eased by visualization in biological networks. Particularly, the reduction of network complexity by visualizing a group of contextually relevant proteins, identification of central nodes and highly interconnected nodes can guide the user in deriving meaning form phosphoproteomics data. However, interpretation of phosphosite information is ultimately more informative for these kinds of studies. Cytoscape is one of the few bioinformatics tools that allows the implementation of quantitative information (such as alteration on phosphosites level), and is therefore particularly advantageous for phosphoproteomics data interpretation.

4 Sequence Bias Analysis

Sequence logos are histogram-like graphical representations of patterns in a list of aligned protein or peptide sequences. The bars in these plots are replaced by letters representing the amino acids at each position. The stack height represents the level of conservation and the height of the individual amino acid reflects their relative frequency at that position [26]. Sequence motif analysis is particularly applicable for phosphoproteomics studies to determine over- and under-represented linear substrate motifs. Many kinases recognize specific sequence motifs. This allows for extrapolation from motif analysis to identify potentially activated kinases and therefore signaling responses in the given biological setting. Most tools to analyze sequence bias also color code the amino acid letter representation based on their physicochemical properties such as hydrophobicity, hydrophilicity, charge, size, and aromatic side chains. Sequence motif analysis should be considered a guide, and hypotheses based on these interpretations require biochemical validations. Here we provide an example of a sequence analysis tool designed for protein sequences, and which provides the option to compare to a reference dataset.

4.1 IceLogo

IceLogo (https://code.google.com/p/icelogo/ or http://iomics.ugent.be/icelogoserver/main.html) [27] is a free, open-source tool for visualization of patterns in aligned protein sequences. This tool can be downloaded or used on a web interface. The advantage of iceLogo compared to other motif analysis tools, is the option to define a custom reference set of sequences. This is particularly important in phosphoproteomics to account for sequence bias in

the phosphopeptide enrichment procedure. Using the reference data, a probability value (Z-score) is calculated, indicating whether the sequences of interest and the reference sequences are similar. If the user cannot provide a reference dataset, the option to choose a reference species of origin is also provided. With this option the frequency of every amino acid in the proteins from the Swiss-Prot database for that species is calculated and used as reference. The results are displayed as position specific bars, heat maps, or so-called iceLogos. The many options to manipulate the visual presentation of the iceLogo can be useful to extract significant information. IceLogo plots are visually easy to interpret, presenting both over- and under-represented patterns. The *X*-axis represents the position in the aligned sequences, and the default *Y*-axis represents percent difference, but can be configured to fold change or standard deviation. The user also has the option to set a *p*-value cutoff. This can be particularly useful to filter noisy sequence motifs, and to find the most significant biases in the sequence. The graphical representation can be saved locally as PDF.

5 Acknowledgments

The authors would like to thank members of the Proteomics Program at the Novo Nordisk Foundation Center for Protein Research (CPR) for critical input on the protocol. Work at CPR is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001).

References

1. Suderman M, Hallett M (2007) Tools for visually exploring biological networks. Bioinformatics 23(20):2651–9
2. Gehlenborg N et al (2010) Visualization of omics data for systems biology. Nat Methods 7(3 Suppl):S56–68
3. da Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37(1):1–13
4. Schnoes AM et al (2013) Biases in the experimental annotations of protein function and their effect on our understanding of protein function space. PLoS Comput Biol 9(5), e1003063
5. Ashburner M et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1): 25–9
6. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
7. Croft D et al (2014) The Reactome pathway knowledgebase. Nucleic Acids Res 42(Database issue):D472–7
8. Dennis G Jr et al (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4(5):P3
9. Finn RD et al (2014) Pfam: the protein families database. Nucleic Acids Res 42(Database issue):D222–30
10. Schultz J et al (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95(11):5857–64
11. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1): 56–68

12. von Stechow L, van de Water B, Danen EH (2013) Unraveling DNA damage response-signaling networks through systems approaches. Arch Toxicol 87(9):1635–48
13. Szklarczyk D et al (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39(Database issue): D561–8
14. Franceschini A et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41(Database issue):D808–15
15. Jensen LJ et al (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37(Database issue):D412–6
16. Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–504
17. Licata L et al (2012) MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 40(Database issue):D857–61
18. Gao J et al (2009) Integrating and annotating the interactome using the MiMI plugin for cytoscape. Bioinformatics 25(1):137–8
19. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2
20. Ideker T et al (2002) Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 18(Suppl 1):S233–40
21. Bindea G et al (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25(8):1091–3
22. Bindea G, Galon J, Mlecnik B (2013) CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 29(5):661–3
23. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21(16): 3448–9
24. Wu G, Feng X, Stein L (2010) A human functional protein interaction network and its application to cancer data analysis. Genome Biol 11(5):R53
25. Wu G, Stein L (2012) A network module-based method for identifying cancer prognostic signatures. Genome Biol 13(12):R112
26. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18(20):6097–100
27. Colaert N et al (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6(11):786–7
28. Lynn DJ et al (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol 4:218
29. Okuda S et al (2008) KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res 36(Web Server issue):W423–6
30. Zeeberg BR et al (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 4(4):R28
31. Hu Z et al (2009) VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. Nucleic Acids Res 37(Web Server issue):W115–21
32. Baitaluk M et al (2006) BiologicalNetworks: visualization and analysis tool for systems biology. Nucleic Acids Res 34(Web Server issue): W466–71
33. Junker BH, Klukas C, Schreiber F (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7:109
34. Kutmon M et al (2013) CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. PLoS One 8(12), e82160

INDEX

Louise von Stechow (ed.), *Phospho-Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1355,
DOI 10.1007/978-1-4939-3049-4, © Springer Science+Business Media New York 2016

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

X

Y

Z