Journal of Chromatography B, 879 (2011) 25–34

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Increasing phosphoproteome coverage and identification of phosphorylation motifs through combination of different HPLC fractionation methods

Xi Chen[a], Di Wu[a], Yong Zhao[a], Barry H.C. Wong[a,b], Lin Guo[a,b,c,*]

[a] College of Life Sciences, Wuhan University, Wuhan, PR China
[b] State Key Laboratory of Virology, Wuhan University, Wuhan, PR China
[c] Key Laboratory of Analytical Chemistry for Biology and Medicine (Ministry of Education), Wuhan University, Wuhan, PR China

ARTICLE INFO

Article history:
Received 3 August 2010
Accepted 2 November 2010
Available online 10 November 2010

Keywords:
HPLC
SCX
HILIC
ERLIC
IMAC
Phosphopeptide

ABSTRACT

Protein phosphorylation activates or deactivates many other proteins especially protein enzymes, and plays a significant role in a wide range of cellular processes. Recent advances in phosphopeptide enrichment procedures and mass spectrometry-based peptide sequencing techniques have enabled us to identify large number of protein phosphorylation sites. In this study, we combined three different HPLC techniques in fractionating enriched phosphopeptides before RPLC–MS/MS analysis, and found that although between 4000–5000 unique phosphopeptides could be identified following any of the HPLC fraction method, different HPLC method yielded a considerable amount of non-overlapping unique phosphopeptides. Combining data from all the HPLC methods, we were able to identify 9069 unique phosphopeptides and 3260 phosphoproteins covering 9463 unique phosphorylation sites, indicating that different HPLC methods are complementary to each other, and can be used together in order to increase the phosphoproteome coverage. A number of new phosphorylation sites and novel phosphorylation motifs were also discovered from our study.

1. Introduction

There are over two hundred different types of post-translational modifications (PTMs), with only a few of those are important for the regulation of biological processes. One of the most studied PTMs to date is protein phosphorylation. Since the first discovery of protein phosphorylation as a regulatory mechanism [1], it has been shown to be a reversible and highly flexible way of influencing the stability, activity and subcellular location of a huge number of proteins [2].

The reversible feature of protein phosphorylation in many proteins is controlled by the combined action of protein kinases and phosphatases, which can maintain stringent tempo-spatial control of phosphorylation and dephosphorylation events [1]. The human genome has revealed that about 2–3% of all genes account for protein kinases and phosphatases, emphasizing the ubiquitous role of protein phosphorylation [3]. It has been estimated that over 50% of all proteins are phosphorylated at some point during its life cycle [4], and that over 100,000 phosphorylation sites may exist in the human proteome [5]. Creating a global network of cellular protein phosphorylation events or phosphoproteome would provide a useful map for better understanding of cellular signal transduction processes.

In recent years, the primary tool for phosphoproteomics has been the use of mass spectrometry [6]. Phosphoproteins are present at relatively low abundance, and phosphorylated forms of individual proteins tend to be present at much lower ratios than their native counterparts [7]. MS-based analysis without phosphopeptide enrichment generally yielded limited results [8]. To overcome this problem, various phosphopeptide enrichment methods such as immobilized metal affinity chromatography (IMAC) and affinity chromatography using metal oxides (such as TiO_2, ZrO_2, etc.) [9,10] have been used to remove nonphosphopeptides from complex peptide mixtures. But if the sample source is derived from cells or tissues, thousands of phosphopeptides can be expected after enrichment step. In order to obtain maximum amount of peptide sequence and phosphorylation site information from these phosphopeptides, pre-fractionation with different LC methods before routine RPLC–MS/MS analysis is an important and necessary step.

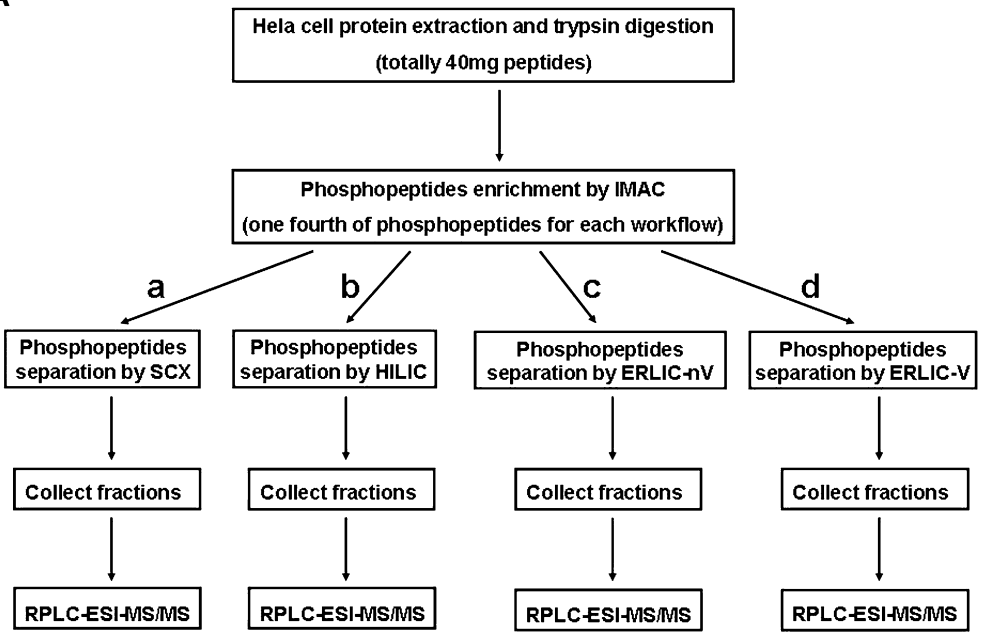
To increase the dynamic range and proteome coverage for bottom-up shotgun proteomics, several effective HPLC methods have been developed for peptide and phosphopeptide fractionation. In SCX (strong cation exchange chromatography), when performed at low pH (pH < 3), tryptic peptides become positively charged by protonation of N-termini and side chains of arginine,

Abbreviations: RP, reverse-phase; MS, mass spectrometry; IMAC, immobilized metal affinity chromatography; SCX, strong cation exchange chromatography; HILIC, hydrophilic interaction chromatography; ERLIC, electrostatic repulsion–hydrophilic interaction chromatography.

* Corresponding author at: College of Life Sciences, Wuhan University, Wuhan 430072, PR China. Tel.: +86 27 68753800; fax: +86 27 68753797.
E-mail address: guol@whu.edu.cn (L. Guo).

1570-0232/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2010.11.004

A



B

Summary of first dimensional HPLC method in separating phosphopeptides

	1st dimension mobile phase A	1st dimension mobile phase B	1st dimension mobile phase C	gradient percent B	gradient percent C
SCX	5mM KH2PO4, 20% ACN, pH2.7	5mM KH2PO4, 20% ACN, 0.5M KCl, pH2.7	*na	0-60	na
HILIC	90% ACN, 0.005% TFA	0.005% TFA	na	0-90	na
ERLIC-nV	20mM Na-MePO3, 70% ACN, pH2.0	200mM TEAP, 60%ACN, pH2.0	na	0-100	na
ERLIC-V	20mM NH4-formate, 70% ACN, pH2.2	20mM NH4-formate, 10% ACN, pH2.2	1M NH4-formate, 10% ACN, pH2.2	0-100	0-100

*na, not applicable.

Fig. 1. (A) Overview of the four different HPLC methods used for phosphopeptide separation. (B) Summary of the different HPLC conditions used. As different dimension HPLC columns were used, different flow rates were adopted accordingly as described in Section 2.

lysine and histidine, while negative charges from carboxyl groups and the C-termini become neutrally charged. Positively charged peptides retained on the SCX resin are then eluted with buffers by increasing the salt concentration or pH. In SCX, peptides are mainly retained according to charge and, in accordance with Coulomb's law, size appears to play a secondary role in the retention explaining the resolution of peptides with the same charge state [11]. SCX is by far the most commonly used and well-established HPLC method for peptides fractionation [12,13], and is often coupled with IMAC (SCX-IMAC) or TiO_2 (SCX-TiO_2) in phosphoproteome analysis [14,15]. HILIC (hydrophilic interaction chromatography) is more commonly used for fractionation of metabolic small polar molecules [16], and less commonly used for peptide fractionation where the primary interaction force between peptides and the neutral hydrophilic stationary phase is hydrogen bonding. Recently, Gilar et al. have shown that HILIC has the highest degree of orthogonality to RPLC of all commonly used peptide separation techniques [17]. Also, Boersema et al. described a zwitterionic HILIC system and used it in a 2D-LC scheme for proteomic applications [18]. In HILIC, retention increases with increasing polarity (hydrophilicity) of peptides. This is opposite the separation principles used in RPLC (reverse-phase liquid chromatography) [19–21]. As hydrophilic and charged phosphopeptides interact more strongly with HILIC than unphosphorylated peptides, it should be possible to separate phosphopeptides by HILIC. Indeed, Zhou's group used HILIC to successfully separate yeast phosphopeptides after IMAC enrichment and identified 8764 unique phosphopeptides from 2278 phosphoproteins [20]. In addition, McNulty and Annan compared two methods including HILIC-IMAC (performing HILIC fractionation before the IMAC enrichment) and IMAC-HILIC (performing HILIC fractionation after the IMAC enrichment) in phosphopeptides fractionation, and found HILIC fractionation dramatically improved the selectivity of IMAC [22]. ERLIC (electrostatic repulsion–hydrophilic interaction chromatography) separation is based on electrostatic repulsion and hydrophilic interactions. It is the newest approach for enriching and fractionating phosphopeptides in a single-step [23].

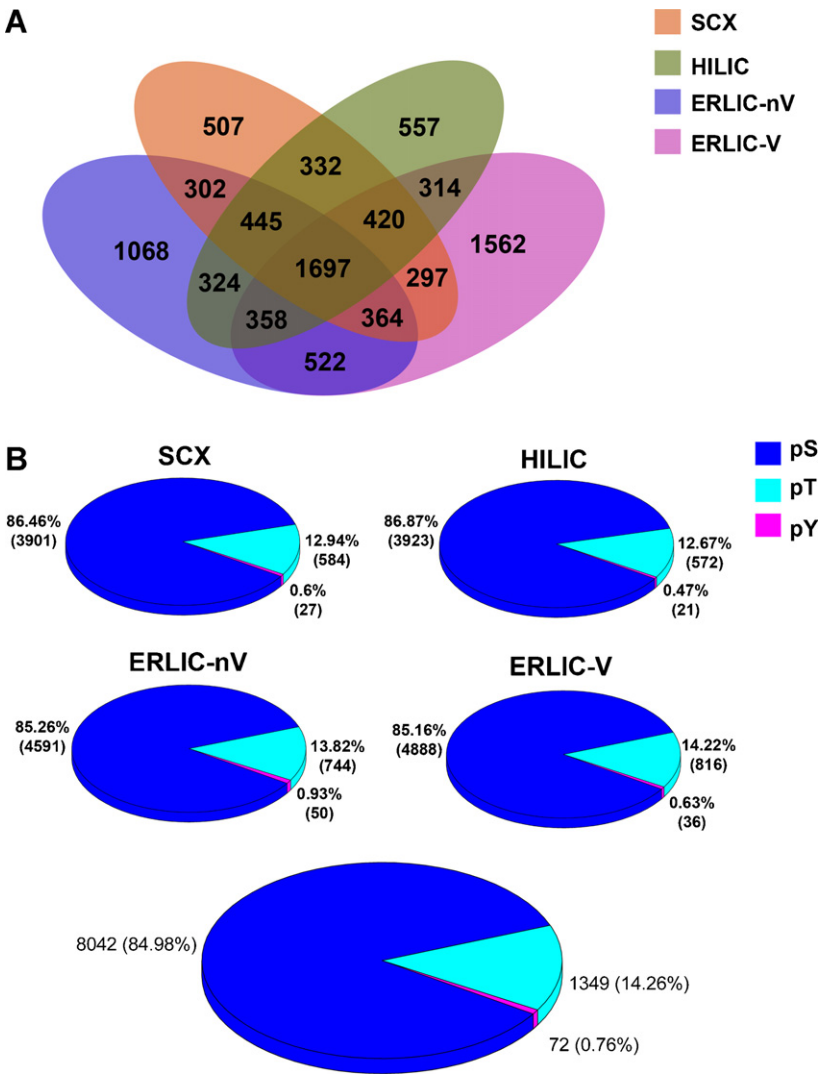


Fig. 2. Venn diagram of overlapping and non-overlapping unique phosphopeptides found from all four HPLC methods (A) and phosphorylation residual distribution (B).

At low pH (pH 2), carboxyl groups at the amino acid residue Asp, Glu, and the C-terminus are uncharged. Phosphate groups retain their negative charge and are electrostatically bound to the column. As this attraction is not sufficient to overcome electrostatic repulsion from the basic amino-acids in a typical tryptic peptide, high concentration of organic solvent (such as 70% acetonitrile) is required to enhance the hydrophilic interaction of the phosphate group within the column [23]. The combination of electrostatic attraction and hydrophilic interaction to separate phosphopeptides from nonphosphopeptides can result in phosphopeptide purification. How well SCX, HILIC and ERLIC perform in phosphopeptide fractionation, and what are their separation characteristics remain to be analyzed.

In this study, for the purpose of increasing phosphoproteome coverage, and to identify novel phosphorylation motifs from the enlarged phosphoproteome data set, we combined SCX-, HILIC- and ERLIC-based HPLC fractionation methods for enriched phosphopeptides separation before sequencing by RPLC–MS/MS. In addi-

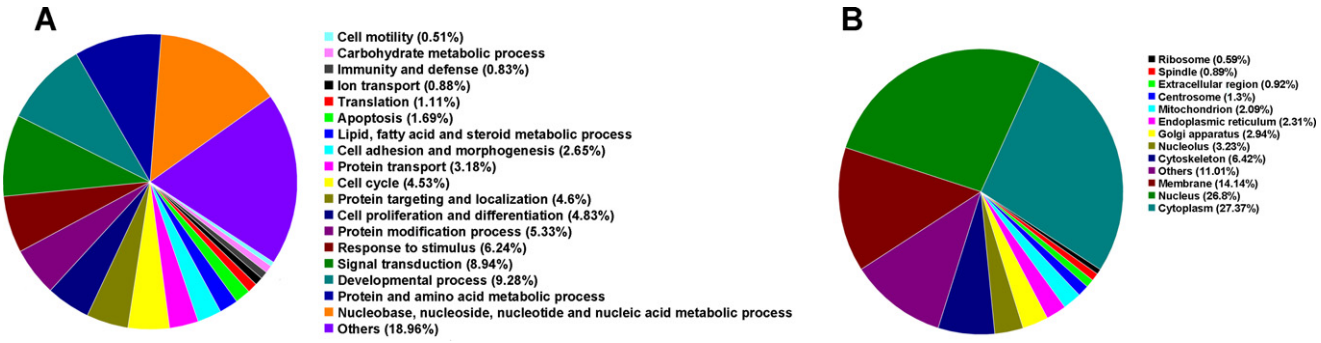


Fig. 3. (A) Functional distribution of all identified phosphoproteins. (B) Subcellular localization of all identified phosphoproteins. All classes not shown have been arranged into the "others" category. The percentage of the total for each class is shown.

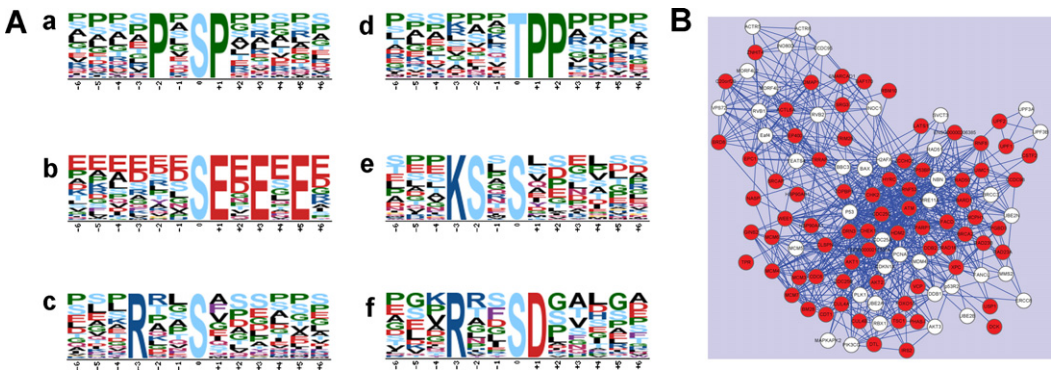


Fig. 4. (A) Sequence logo representations of single-phosphorylated motifs where the serine or threonine is phosphorylated: (a) proline-directed consensus sequences of GSK-3, ERK1, ERK2 and CDK5 kinase substrates; (b) 76 sequences (motif score = 44.31) were found to contain the acidic motif from our data set; (c) basic motif representative of a calmodulin-dependent protein kinase II kinase substrate; (d) a proline-directed motif containing a phosphothreonine with two proline residues adjacent to the phosphate; (e) basic motifs with high scores but lacking any known protein kinase phosphorylation site; (f) a special motif that contains a basic residue N-terminal to the pSer and an acidic residue C-terminal to the pSer. (B) Protein kinases and substrates networks in response to DNA damage. Phosphoproteins identified in our study set are labeled in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

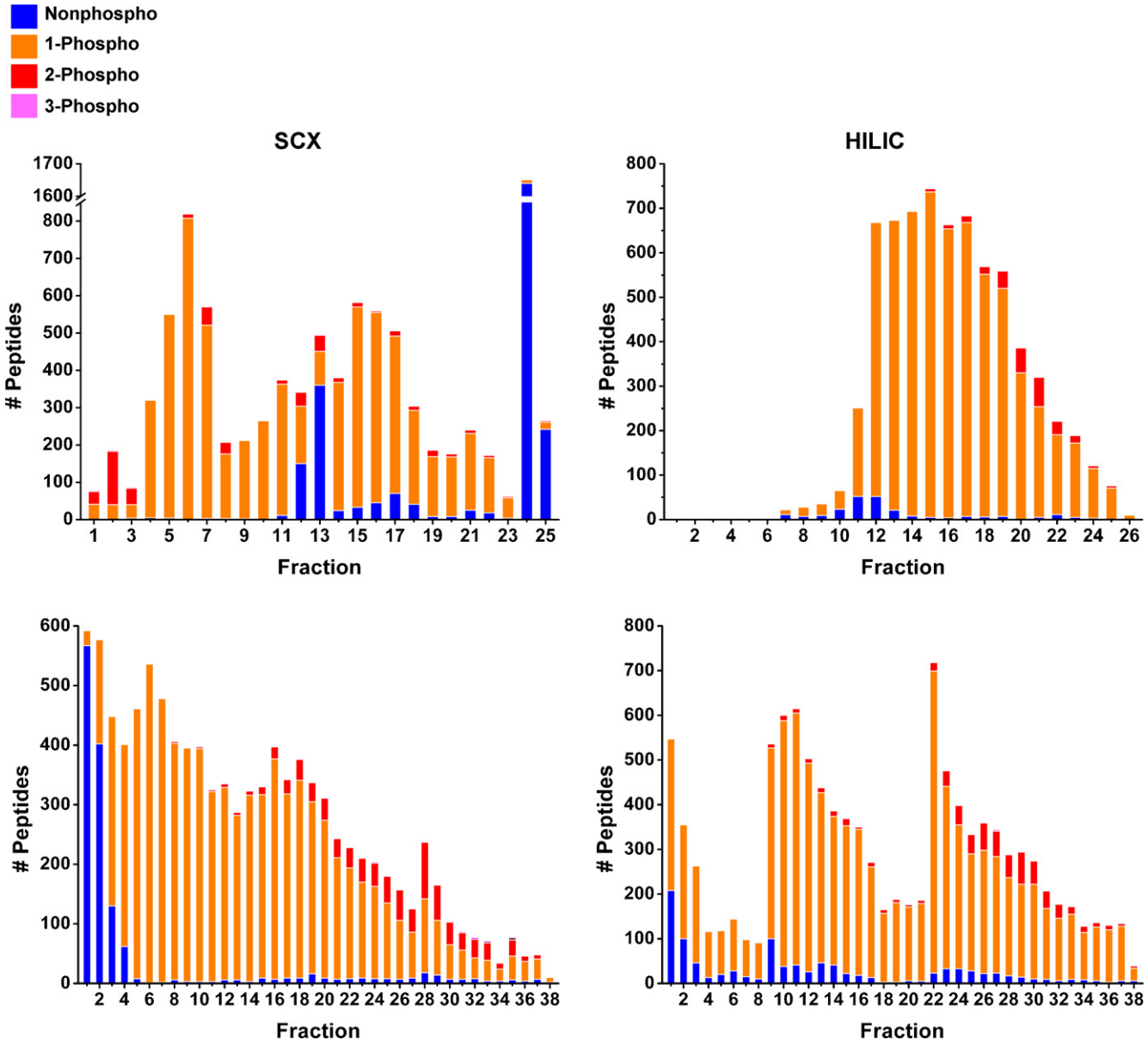


Fig. 5. Distribution of nonphosphopeptides and phosphopeptides with one, two or three phosphorylation sites obtained by different HPLC methods.

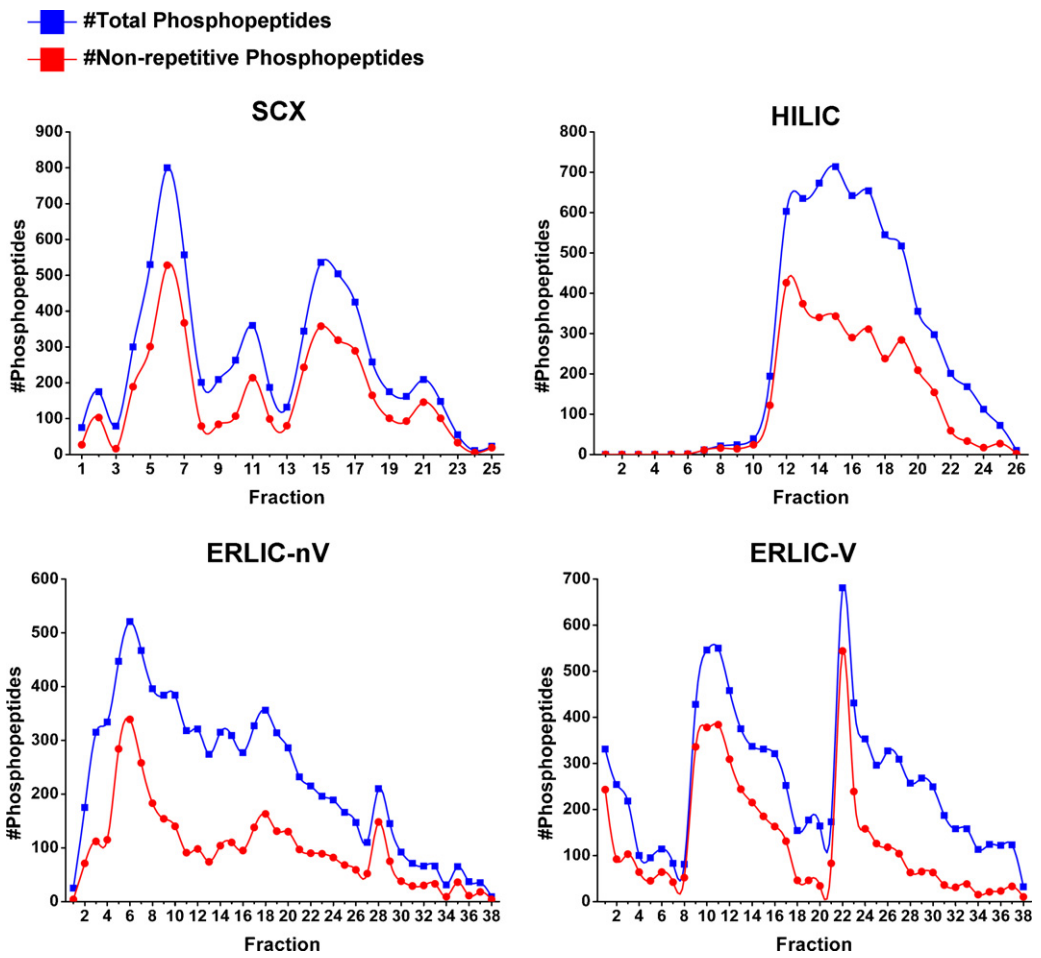


Fig. 6. Comparison of phosphopeptide resolution from different HPLC methods. The total phosphopeptides represent phosphopeptides from every fraction while non-repetitive phosphopeptides represent phosphopeptides found in only one fraction between three adjacent fractions. The blue curve represents cumulative total number of phosphopeptides while the red curves represents the number of non-repetitive phosphopeptides in each fraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

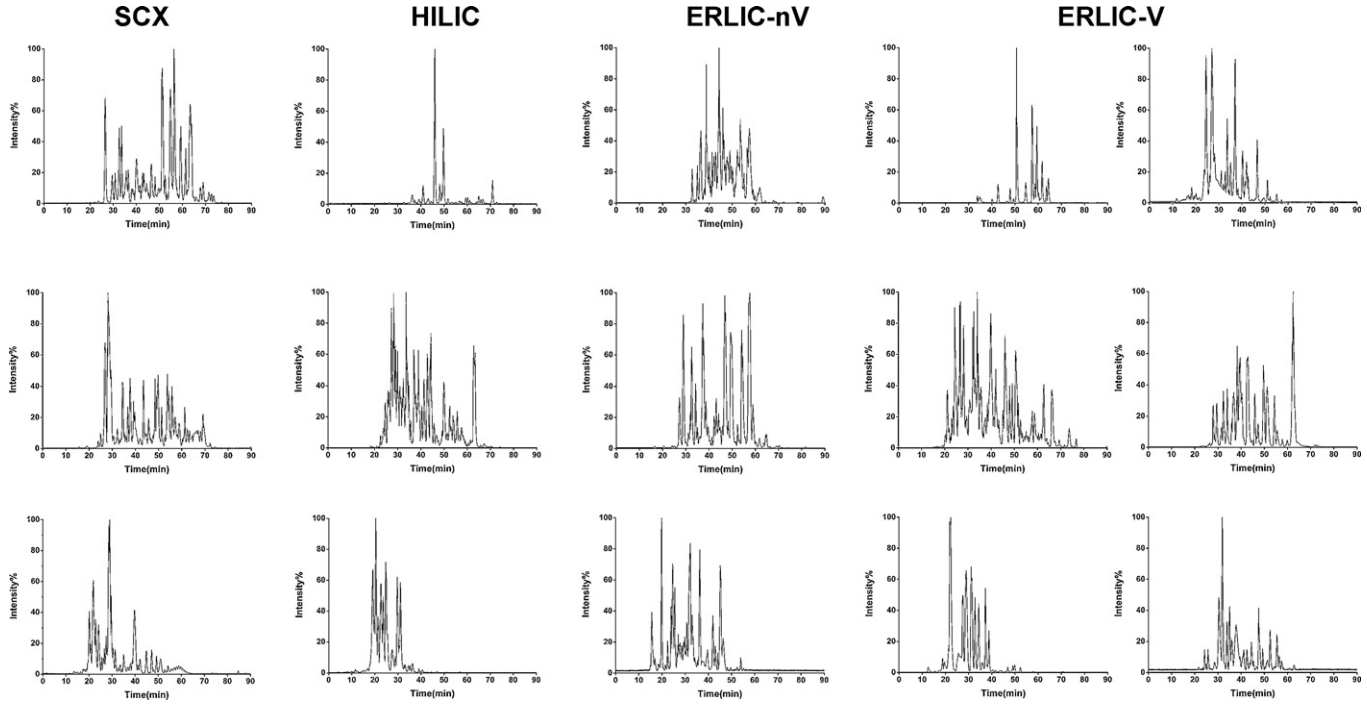


Fig. 7. Comparison of orthogonality from different HPLC methods. Representative base peak chromatograms (BPC) of fractions from the beginning, middle, and end of SCX, HILIC, ERLIC-nV and ERLIC-V HPLC.

tion, we systematically compared those HPLC methods mentioned above in phosphopeptide fractionation. In order to achieve this aim, phosphopeptides and nonphosphopeptides distribution, multi-phosphopeptides distribution, as well as various chromatographic factors that impact HPLC separation efficiencies were compared.

2. Materials and methods

2.1. Cell culture, protein extraction and digestion

HeLa cell lines were obtained from the ATCC. Cells were cultured in DMEM supplemented with 10% new born calf serum (GIBCO, Invitrogen Inc., Waltham, MA), 100 U of penicillin, and 100 μg of streptomycin per mL (Invitrogen, Carlsbad, CA). The cells were maintained at 37 °C in a 5% CO_2 atmosphere. When the cell density reached approximately 90% confluence, the cells were lysed in buffer containing 25 mM Tris–HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, and 0.1% SDS. Phosphatase inhibitor (PhosSTOP, Roche Applied Science) and protease inhibitor cocktails (COMPLETE tablets, Roche Applied Science) were added at the recommended doses. To precipitate proteins, 3 volumes of 50% acetone/50% ethanol/0.1% acetic acid was added to cell lysates and placed on ice for 1 h. The sample was centrifuged at 4000 rpm (4 °C) for 20 min, and the protein pellet was re-suspended in 8 M urea, 0.2 M Tris (pH 8), 4 mM $CaCl_2$, and reduced in 10 mM DTT for 1 h at 56 °C. The sample was cooled to room temperature and alkylated by incubating in 40 mM iodoacetamide for 30 min in the dark. After diluting the sample in 7–8 volumes of H_2O, the final protein concentration was measured and trypsin was added to a final ratio of 1:50 trypsin/protein (w/w). The samples were then digested by incubating on a rocking shaker overnight at 37 °C. The digested sample was desalted by loading on to two 2 g Sep-Pak C18 columns (Waters), washed twice with 10 mL of 1% acetic acid, and eluted in 7 mL of 80% acetonitrile and 0.1% acetic acid. The eluate from each column was speed-vac dried and stored at −80 °C until use.

2.2. Phosphopeptide enrichment.

Phosphopeptides were enriched from tryptically digested protein mixtures generated from 40 mg cell lysate. Phosphopeptide enrichment was performed according to a published protocol [20].

Dried peptides were re-suspended in 400 μL of 1% acetic acid and loaded onto 8 gel loading tip columns with each tip containing 40 μL of IMAC resin. After loading, the IMAC resin was washed twice with 40 μL of wash buffer containing 25% acetonitrile, 100 mM NaCl and 0.1% acetic acid, followed by single washes with 40 μL of 1% acetic acid and 20 μL of deionized water. Phosphopeptides were eluted from the tip with 120 μL of 6% NH_4OH and dried under vacuum.

2.3. Phosphopeptide separation

Phosphopeptides from a single batch preparation was used. The first dimensional HPLC was performed using an Agilent 1200 system (Agilent Technologies). The HPLC mobile phase and separation gradient of each workflow are listed in Fig. 1B.

For SCX, a PolySULFOETHYL A column (2.1 mm × 50 mm, 5 μm particle size, 200 Å pore size) (PolyLC, Columbia, MD) was used, and UV detection was monitored at a wavelength of 216 nm. A total of 25 fractions were collected by collecting fractions at 2 min intervals. For HILIC, a TSK gel Amide-80 column (2.0 mm × 150 mm, 5 μm particle size, 200 Å pore size) (TOSOH Bioscience) was used, and UV detection was monitored at a wavelength of 215 nm. A total of 26 fractions were collected by pooling fractions at 2 min intervals. Collected fractions from SCX and HILIC were dried under a vacuum. For ERLIC separation, A PolyWAX LP column (4.6 mm × 200 mm, 5 μm particle size, 300 Å pore size) (PolyLC) was used, and UV detection was monitored at a wavelength of 214 nm. Two separate solvent systems were utilized, one using a non-volatile reagent, and the other using volatile reagent (Fig. 1B).

All fractions that were separated by SCX and ERLIC-nV were desalted by ZiptipC18 (Millipore) and performed according to the manufacturers' instructions.

2.4. RPLC–ESI-MS/MS

A nano flow multiple dimensional HPLC system (Tempo™ nano MDLC system, Applied Biosystems) was coupled online with a QSTAR ELITE mass spectrometer (Applied Biosystems) and used for RPLC–ESI-MS/MS analysis. Peptides were first trapped within a CapTrap column (0.5 mm × 2 mm, MICHROM Bioresources, Inc.) and eluted into an integrated nanoscale analytical column MAGIC C18AQ (100 μm × 150 mm, 3 μm particle size, 200 Å pore size, MICHROM Bioresources, Inc.). Mobile phase A (2% ACN, 0.1% formic acid) and mobile phase B (98% ACN, 0.1% formic acid) were used to establish a 130 min gradient, which comprised of: 5 min in 5% B, 25 min of 5–15% B, 55 min of 15–40% B, 15 min of 40–80% B, the gradient was maintained in 80% B for 10 min, followed by 5 min of 80–5% B, and a final step in 5% B for 15 min. A constant flow rate was set at ∼300 nL/min. MS scans were conducted from 400 to 1800 amu, with a 1 s time span. For MS/MS analysis, each scan cycle consisted of one full-scan mass spectrum (with m/z ranging from 400 to 1800 and charge states from 2 to 5) followed by five MS/MS events. The threshold count was set to 30 to activate MS/MS accumulation and former target ion exclusion was set for 90 s. The mass tolerance was set at 50 mDa, and Automatic Collision Energy and Automatic MS/MS Accumulation were selected.

2.5. Data analysis

Raw data from QSTAR ELITE were analyzed with Mascot Daemon software (version 2.2.2) (Matrix Science, London, UK) using an in-house MASCOT server (version 2.2) (Matrix Science, London, UK). Data was searched against the SwissProt human protein database (version 56.9; 20402 sequences) using the following parameters: Fixed modifications was set to carbamidomethylation on cysteine, variable modifications was set to oxidation of methionine, and phosphorylation at serine, threonine or tyrosine, and the taxonomy was set to "human". Peptide and MS/MS tolerances were set at 50 ppm and 0.2 Da, respectively. The peptide charge was set to 2+, 3+, 4+, or 5+, allowing for up to two missed cleavages, and the significance threshold was set at $p < 0.05$. After the MASCOT search, the .dat data was imported into Scaffold (version 2_04_00) for further analysis. In order to select phosphopeptides with high confidence, "Min Protein", "Min # Peptide" and "Min Peptide" were adjusted to 20%, 1 and 95% respectively. The .dat data was further analyzed with the X!Tandem search engine and was integrated into Scaffold.

2.6. Motif analysis

Possible phosphorylation motifs present in our data set were obtained with the Motif-X algorithm [24]. Phosphorylated residues are denoted in lowercase letters, and non-phosphorylated residues as uppercase letters. Peptide sequences were centered on each phosphorylation site, and extended to 13 amino acids in length. The IPI human database was set as the background. The minimum number of occurrences was set to 20, and the significance threshold was set to 10^{-6}. The major motif classes were defined and based on the composition of the amino acid residues. The identified classes were

Table 1
Comparison of four HPLC methods used in phosphopeptides separation.

	SCX	HILIC	ERLIC-nV	ERLIC-V
Capability in separation between phospho- and nonphospho-peptides after IMAC	Good	Low	Good	Medium
Capability in separation between mono- and multi-phosphopeptides	Medium	Low	Excellent	Good
Resolution in phosphopeptides separation	Good	Low	Low	Medium
Orthogonality relative to RP	Good	Low	Good	Medium
Whether or not to need desalting	Yes	No	Yes	No

proline-directed (P containing sequences), acidic (D and E containing sequences) and basic (R and K containing sequences) motifs [6]. The identified motifs from our experiments were compared by cross-referencing with the Human Protein Reference Database [25] for known kinase substrates.

2.7. Phosphoprotein network analysis

Phosphorylated proteins involved in the DNA damage response (DDR) were summated and analyzed by STRING [26] so that the protein IDs (nodes) and protein–protein interactions (edges) can be extracted. The obtained network was then loaded onto Cytoscape (version 2.6.3) [27] for data visualization. Interactions with a minimum STRING score of 0.400, which represents the default medium confidence level in STRING, were kept for analysis.

3. Results and discussion

3.1. Phosphopeptide identification.

Our experimental workflow is shown in Fig. 1A. A summary of the HPLC conditions tested are listed in Fig. 1B.

Phosphopeptide identification was based on MASCOT program analysis with >95% confidence setting. The false discovery rates (FDRs) at the protein level as calculated by Scaffold were 0.74% (SCX), 0.51% (HILIC), 0.39% (ERLIC-nV) and 0.43% (ERLIC-V), respectively. The total number of unique phosphopeptides identified (with >95% confidence) from the 4 methods ranged from 4000 to 5000 (Fig. 2A). This is significantly higher than the number of phosphopeptides identified from a single RPLC–MS/MS run without any prior fractionation (~700 unique phosphopeptides per run). A total of 9069 unique phosphopeptides were identified when data combined (Fig. 2A, detailed phosphopeptide list is in supplementary Table S1). In addition, only 1697 unique phosphopeptides (18.7% of total unique phosphopeptides) were found in all four HPLC methods (Fig. 2A). There are many non-overlapping unique phosphopeptides in any two methods and the lowest ratio of non-overlapping unique phosphopeptides in total unique phosphopeptides is 51% (Fig. 2A), indicating that different HPLC methods can be used in complementary to achieve a more complete phosphoproteome. We infer that this phenomenon of high non-overlapping ratios is relative to different separation principles of four HPLC methods. It is supposed that a specific peptide with low abundance appears in flow-through by one HPLC method and there are thousands of kinds of peptides in this fraction. So this peptide cannot be detected by MS because of ionization suppression by other peptides. However, in another HPLC method, this peptide may be eluted into a fraction which contains a few kinds of peptides. Then, this peptide can be detected without interference.

The distribution of phosphorylated amino acid residues on phosphopeptides was similar in all four HPLC methods with the most phosphorylation occurring on Ser followed by Thr and Tyr (Fig. 2B), which is consistent with previously reported studies [6,15]. In total, we identified 9463 unique phospho-sites by combining data from all four HPLC methods, and the distribution of phosphorylated amino acid residues was shown in Fig. 2B. In a previous report, Mitulovic et al. used ERLIC to enrich and fractionate phosphopeptides from tryptic digest of HeLa proteins and obtained a relative high pTyr distribution (~6%) [28], suggesting that ERLIC can identify more phosphopeptides of low abundance. In our study, the pTyr distribution from results by ERLIC-nV method was ~0.93% (Fig. 2B), much lower than the 6% observed by Mitulovic et al. This discrepancy may be caused by IMAC enrichment of phosphopeptides before HPLC separation.

3.2. Gene ontology analysis

In the combined data set, a total of 3260 phosphoproteins with FDR < 1% were identified after Scaffold analysis of the 9069 phosphopeptides. In order to illustrate the broad phosphoproteome coverage, the 3260 phosphoproteins identified were annotated for GO biological process and cell component. In "biological process" terms, these phosphoproteins were mainly involved in nucleotide and nucleic acid metabolic process, protein and amino acid metabolic process, developmental process, signal transduction, response to stimulus, protein modification process, cell proliferation and differentiation, protein targeting and localization, cell cycle, protein transport, etc. (Fig. 3A), which indicated a comprehensive functional distribution. In "cell component" terms, most phosphoproteins were involved in cytosol, nucleus, membrane, cytoskeleton, nucleolus, Golgi apparatus, endoplasmic reticulum, mitochondrion, etc. (Fig. 3B), which showed extensive cellular localizations.

3.3. Phosphorylation motif and network coverage analysis

The 9069 unique phosphopeptides identified were subjected to phosphorylation motif analysis with the Motif-X algorithm [24] resulting in assignment of specific and frequency-corrected phosphorylation motifs to the phosphopeptides (Fig. 4A). In total, we identified 53 motifs relating to pSer and 12 motifs relating to pThr (see supporting information, Table S2). No motifs were identified for pTyr due to the small sample size. The identified pSer and pThr motifs were classified into 3 main categories: Proline-directed motifs (38%), acidic motifs (38%), and basic motifs (12%).

Logo-like representations were created to display each identified motif graphically. These consensus sequences not only include residues that strictly adhere to the motif (see supporting information, Table S2) but also represent the frequencies of adjacent amino acids observed and identified around the phosphosite (Fig. 4A). Several motifs identified were found to conform to known kinase substrates. For example, sEXEXE, sDEE and sDXE are all associated with Casein Kinase II substrates. We also identified a number of potentially 'new' motifs, for example the tPP motif (with 227 occurrences) and the KSXs motif (with 227 occurrences), which have not been reported to be substrates of any particular kinase. A unique motif, RxxsD, which contains a basic residue on the N-terminal side of the pS and an acidic residue on the C-terminal of the pS was identified. The RxxsD has the potential to be phosphorylated by both acidiphilic and basophilic kinases.

DNA damage response (DDR) is a global signaling network that senses different types of DNA damage or replication stress, which can result in a multitude of responses which include: transcriptional activation, cell cycle control, apoptosis, senescence and DNA repair processes. DDR is critical for cell survival and has important implications in aging and cancer [29]. At the core of DNA damage signaling apparatus, a series of related protein kinases and substrates play important roles. Major phosphorylated proteins that have been known to participate in the DDR pathway [29,30] were picked out and a protein–protein interaction network was constructed with STRING and Cytoscape software. Among the 116 published phosphoproteins involved in DDR, 75 phosphoproteins (~65%) were identified from our combined data set (Fig. 4B). Low abundant phosphoproteins such as BRCA1-associated RING domain protein 1 (BARD1), E3 ubiquitin-protein ligase Mdm2 (Mdm2) were also detected. In addition, we identified pS357 in Chk1 and pS379 in Chk2, which have not been previously known to be phosphorylation sites. In Chk1, pS357 site is adjacent to a residue Q which suggests that this site may be phosphorylated by the ATM/ATR protein kinases. Thirty-two phosphorylation sites were identified in another important DDR regulated phosphoprotein, Mdc1 (Mediator of DNA damage checkpoint protein 1), with ten of the phosphorylation sites (S485, T1164, T1201, S1441, T1447, T1485, T1488, S1563, S1570 and T1800) previously unpublished.

3.4. HPLC separation characteristics analysis

Since little literature information available that focus on HPLC separation of complex phosphopeptide mixtures, we compared the separation characteristics of the four different HPLC methods, and tried to understand how different HPLC methods separated phosphopeptides from the aspects of selectivity, resolving power and orthogonality.

3.4.1. Phosphopeptides distribution in different HPLC methods

The distribution of phosphopeptides from the SCX separation was that: the early fractions contained many phosphopeptides derived from the protein C-terminus without either terminal arginine or lysine residue, also a lot of phosphopeptides containing acid amino-acid residues (D or E) appeared in the early fractions, whereas phosphopeptides from late fractions were larger and basic often with missed tryptic cleavages. The distribution of phosphopeptides from the HILIC separation was, as expected, based on hydrophilicity. The early fractions were largely comprised of short, hydrophobic phosphopeptides, whereas phosphopeptides from late fractions were larger and hydrophilic often with acid amino-acid residues (D or E) and missed tryptic cleavages. The distribution of phosphopeptides from the ERLIC-nV separation was also consistent with its principle. The early fractions were largely comprised of short, hydrophobic phosphopeptides often with missed tryptic cleavages, whereas phosphopeptides from late fractions were larger and hydrophilic often with acid amino-acid residues (D or E). The distribution of phosphopeptides from the ERLIC-V separation was irregular.

3.4.2. Comparison of phosphopeptides and nonphosphopeptides distribution

A large variation in distribution of the identified unique nonphosphopeptides and phosphopeptides from a single RPLC–MS/MS was observed (Fig. 5). Out of the 4 HPLC separation schemes tested, SCX and ERLIC-nV provided greater separation between phospho- and nonphosphopeptides, while HILIC method had the lowest ability and could only identify 174 unique nonphosphopeptides (Fig. 5). Most of the nonphosphopeptides isolated composed of peptides rich in acidic amino-acid residues (D or E) (data not shown). It is well known that one of the problems in IMAC is the high level of nonspecific binding when used for phosphopeptide enrichment of highly complex peptide samples. Most of the nonphosphopeptides contain multiple acidic amino-acid residues (D or E) and copurify with the phosphopeptides, resulting in a serious problem of suppression of signals from phosphopeptides caused by more efficient ionization of nonphosphopeptides in subsequent MS analysis. In order to circumvent this problem, Ficarro et al. designed an experiment to derivatize the carboxylic groups on acidic amino acid residues in peptides by O-methyl esterification [9] and thereby considerably improved the efficiency of phosphopeptides enrichment by IMAC. But further studies show that O-methyl esterification cannot derivatize all the carboxylic acid groups in 100% efficiency [31] and often causes a partially deamidation and subsequent methylation of Asn and Gln residues [10] which will increase the complexity of MS analysis and data interpretation due to the peptides with different degrees of O-methylation and byproducts. In our study, we can solve this intractable problem in IMAC used for phosphopeptides enrichment of highly complex peptide samples by further separation of phospho- and nonphospho-peptides through SCX or ERLIC-nV HPLC method without any extra chemical derivatization. SCX or ERLIC-nV HPLC separation can be regarded as a second phosphopeptide enrichment step following IMAC phosphopeptide enrichment, which can further enhance the purity of phosphopeptides in most collected fractions and then increase the number of identified phosphopeptides because of the reduction of ionization suppression caused by nonphosphopeptides.

3.4.3. Comparison of multiphosphopeptides distribution

The number of double or triple phosphorylated peptides identified varied significantly for all four HPLC methods (SCX, HILIC, ERLIC-nV, ERLIC-V) (Fig. 5). For double phosphorylated peptides, the numbers identified from these four methods ranged from 186 to 694. ERLIC-nV method identified the most unique double phosphorylated peptides. These double phosphorylated peptides were primarily coming out of late fractions (Fig. 5). The ERLIC-V method followed next because the eluting salt used in ERLIC-V was mild compare to TEAP in ERLIC-nV method (TEAP could elute almost all multiple phosphorylated peptides) [32]. The best result in separating mono- and multi-phosphopeptides was achieved in ERLIC methods, which was consistent with ERLIC separation principle: the positively charged matrix in WAX column binds more firmly with higher numbers of negatively charged phosphate groups, therefore enriching multiphosphopeptides in the late phase of chromatography [23]. In SCX, double phosphorylated peptides were mainly concentrated in the initial three fractions with a total of 351 double phosphorylated peptides identified from all SCX fractions (Fig. 5). The low number of multiple phosphorylated peptides detected by SCX may be due to the formation of weak interactions of peptides containing multiple negatively charged phosphate groups with the negatively charged matrix on the SCX column [33], and we suspected that some multiphosphopeptides were lost in the flow-through. The HILIC method yielded the least unique double phosphorylated peptides, indicating low resolving power for mono- and multi-phosphopeptides.

Although ERLIC-nV method could identify the most double phosphorylated peptides, the number of double phosphorylated peptides identified was still much lower than the monophosphopeptides (694 vs 4386). Moreover, the number of identified triple phosphorylated peptides was low. This phenomenon was inconsistent with some previous reports [28,34]. The reason can be explained by (1) phosphopeptide enrichment conditions prior to HPLC separation. During IMAC, the sample was dissolved in an acetic acid solvent producing a final pH 3–4, which was favorable for monophosphopeptide enrichment. If the dissolving solvent was lowered to pH 2–3, a higher proportion of multiphosphopeptides could be enriched (data not shown). (2) The use of additional

desalting steps. With the exception of HILIC, all fractions collected from the other three HPLC methods were desalted by ZipTip before RPLC–MS/MS analysis. Multi-phosphorylated peptides are more hydrophilic and bind less tightly to reverse-phase C8 or C18 beads particularly in the presence of salt, making the loss of mutiphosphopeptides unavoidable [20]. (3) The limitation on instrument performance and scan rate, resulting in the loss of identification of the less abundant mutiphosphopeptides [35]. (4) Collision-induced dissociation (CID) fragmentation of multiphosphopeptides usually results in the loss of phosphoric acid with poor peptide backbone fragmentation, and consequently little sequence information is obtained. However electron capture/transfer dissociation (ECD/ETD) can give better results for mutiphosphopeptides [36,37].

3.4.4. Comparison of resolution in phosphopeptides separation

During a HPLC separation, it is common for a single peptide to appear in two or more adjacent fractions, and this is reflected in the resolution of method. If we define a non-repetitive phosphopeptide as one that only appears in one fraction and count the number of non-repetitive and repetitive phosphopeptides in every fraction, a resolution plot of the 4 HPLC methods can be produced (Fig. 6). In the plot, blue curves represent the cumulative total number of phosphopeptides, while the red curves represent the number of non-repetitive phosphopeptides in each fraction. A large gap between the red curve and the blue curve is indicative of poor resolution. The SCX method showed high degree of coincidence between the red and the blue curve (Fig. 6). When numbers of peptides from all the fractions were summated, the ratios of non-repetitive phosphopeptides to total phosphopeptides for SCX, HILIC, ERLIC-nV, ERLIC-V methods were 60.52%, 50.79%, 43.61%, and 50.83%, respectively. Our overall results show that SCX has the best resolving power for phosphopeptides separation. The use of longer SCX columns would offer better separation/resolution. As previous reported [33], phosphopeptides mainly existed in flow-through and front fractions because of their negatively charged phosphate group in SCX chromatography. But in our experiments, what is noteworthy is that phosphopeptides are evenly distributed throughout SCX fractions. We speculate that in SCX, phosphopeptides are mainly retained according to charge, and phosphopeptide size may play another important role in the retention of some phosphopeptides with the same charge state [11], which would contribute to the overall separation power of SCX for phosphopeptides. In HILIC, phosphopeptides mainly distribute in the middle fractions. This was consistent with the results reported by McNulty and Annan [22].

3.4.5. Orthogonality comparison of four HPLC methods

In addition to resolution, the orthogonality of separations is another factor that can affect and contribute to the overall separation efficiency. Orthogonality reflects the relative difference in the selectivity between the first and second dimensional separation modes [38]. The orthogonality of all four HPLC methods was compared by measuring the distribution of phosphopeptides during second dimensional RP-HPLC (Fig. 7). The chromatograms were obtained by base peak chromatography (BPC) of eluted peptides, and consisted of three or six discrete fractions obtained from the beginning (not flow-through), middle, and end of first dimensional separation (SCX, HILIC, ERLIC-nV and ERLIC-V). During SCX, the phosphopeptides are evenly distributed in all three chromatograms, denoting good orthogonality. This is not unexpected as selectivities of SCX and reverse-phase are very different [38]. In contrast, the HILIC method produced a staggered elution profile from the early to the late fraction. The first fraction exhibited an obvious retention time bias towards late elution, the middle fraction showed no bias, and the last fraction exhibited an obvious retention time bias towards early elution. This profile is due to HILIC utilizing hydrophilic interactions for separation while RP utilizes hydrophobic interactions to separate peptides. These two methods utilize opposing separation principles. Phosphopeptides from the initial fractions in HILIC appear in the late elution of RP and phosphopeptides of later fractions from HILIC appear in the early elution of RP, resulting in semi-orthogonal separation. In the ERLIC-nV method, a mix of hydrophilic interaction and ion exchange produces different selectivity from RP, resulting in good orthogonality. In ERLIC-V, we used two different elution conditions: during first phase elution, the HPLC conditions exhibited a profile similar to HILIC whereby the orthogonality was not ideal; in second phase elution, the HPLC condition was similar to SCX, where higher orthogonality was observed.

3.5. Reproducibility analysis

We analyzed a complex phosphopeptide sample in parallel using RPLC–ESI-MS/MS without fractionation and found approximately 85% peptides were overlapping. Of course, the phosphopeptide profile obtained from an individual RPLC–ESI-MS/MS may be influenced by the conditions of the HPLC column, and by the mass spectrometer instrument. However, it is reasonable to believe that the complementary feature of the different HPLC separation methods should be reproducible, because it was quite clear that the profiles of different HPLC methods in phosphopeptide separation reflected the method's separation principles (Fig. 5).

4. Conclusion

In our study we adopted four different HPLC methods coupled with RPLC MS/MS analysis to obtain a very comprehensive coverage of phosphoproteome, and found that each HPLC method has its strengths and weaknesses (Table 1). SCX, the classical method for mixed peptide fractionation showed good resolution and orthogonality in separating phosphopeptides, however an additional desalting step is required prior to RPLC–MS/MS analysis, not only making the method inconvenient but also increasing the risk of phosphopeptide (especially multi-phosphopeptides) loss. But recently Andrew's group discovered that different fractions (high-salt or low-salt fractions) desalted by different materials could improve recovery of phosphopeptides [39]. Although the resolution and orthogonality of the HILIC method is not ideal, one of the key advantages of HILIC is the lack of a desalting step before RPLC–MS/MS, which can cut costs and time. The ERLIC-nV method provided the highest number of phosphopeptides containing two phosphorylation sites, good separation between nonphospho- and phospho-peptides after IMAC, and good orthogonality with reverse-phase separation before MS analysis. In our ERLIC-V experiment, although the fractionated samples were desalted before RPLC–MS/MS, this step might not be necessary as the volatile salt from ERLIC-V could be removed by re-solubilizing in small volumes of water:methanol (1:1) and drying 2–3 times prior to MS analysis [39]. All methods were able to identify large numbers of unique phosphopeptides with each method yielding a considerable amount of non-overlapping unique phosphopeptides because of different HPLC separation characteristics. In the combined data set, more than half of the known phosphoproteins from the DNA damage response network were found and some unpublished phosphorylation sites and novel phosphorylation motifs were identified, demonstrating the power of combining phosphopeptide enrichment with different effective fractionation methods.

Acknowledgements

This work was supported by the National Basic Research Program of China (2007CB914200), the National Science Foundation of China (#30921001 and #30950110327), and the PCSIRT (IRT0745).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2010.11.004.

References

[1] E.G. Krebs, J.A. Beavo, Annu. Rev. Biochem. 48 (1979) 923.
[2] J.D. Graves, E.G. Krebs, Pharmacol. Ther. 82 (1999) 111.
[3] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., Science 291 (2001) 1304.
[4] J. Reinders, A. Sickmann, Proteomics 5 (2005) 4052.
[5] D.E. Kalume, H. Molina, A. Pandey, Curr. Opin. Chem. Biol. 7 (2003) 64.
[6] J. Villen, S.A. Beausoleil, S.A. Gerber, S.P. Gygi, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 1488.
[7] T.E. Thingholm, O.N. Jensen, M.R. Larsen, Proteomics 9 (2009) 1451.
[8] A. Schmidt, E. Csaszar, G. Ammerer, K. Mechtler, Proteomics 8 (2008) 4577.
[9] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nat. Biotechnol. 20 (2002) 301.
[10] M.R. Larsen, T.E. Thingholm, O.N. Jensen, P. Roepstorff, T.J. Jorgensen, Mol. Cell. Proteomics 4 (2005) 873.
[11] K. Sandra, M. Moshir, F. D'Hondt, R. Tuytten, K. Verleysen, K. Kas, I. Francois, P. Sandra, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 877 (2009) 1019.
[12] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.
[13] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Nat. Biotechnol. 19 (2001) 242.
[14] J. Villen, S.P. Gygi, Nat. Protoc. 3 (2008) 1630.
[15] J.V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Cell 127 (2006) 635.
[16] H. Idborg, L. Zamani, P.O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 828 (2005) 9.
[17] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Anal. Chem. 77 (2005) 6426.
[18] P.J. Boersema, N. Divecha, A.J. Heck, S. Mohammed, J. Proteome Res. 6 (2007) 937.
[19] A.J. Alpert, J. Chromatogr. 499 (1990) 177.
[20] C.P. Albuquerque, M.B. Smolka, S.H. Payne, V. Bafna, J. Eng, H. Zhou, Mol. Cell. Proteomics 7 (2008) 1389.
[21] D.E. McNulty, R.S. Annan, Methods Mol. Biol. 527 (2009) 93.
[22] D.E. McNulty, R.S. Annan, Mol. Cell. Proteomics 7 (2008) 971.
[23] A.J. Alpert, Anal. Chem. 80 (2008) 62.
[24] D. Schwartz, S.P. Gygi, Nat. Biotechnol. 23 (2005) 1391.
[25] R. Amanchy, B. Periaswamy, S. Mathivanan, R. Reddy, S.G. Tattikota, A. Pandey, Nat. Biotechnol. 25 (2007) 285.
[26] L.J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, C. von Mering, Nucleic Acids Res. 37 (2009) D412.
[27] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Genome Res. 13 (2003) 2498.
[28] G. Mitulovic, A.J. Alpert, K. Mechtler, poster presented at 2008 ASMS, Meeting (2008).
[29] J.W. Harper, S.J. Elledge, Mol. Cell 28 (2007) 739.
[30] S. Matsuoka, B.A. Ballif, A. Smogorzewska, E.R. McDonald 3rd, K.E. Hurov, J. Luo, C.E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S.P. Gygi, S.J. Elledge, Science 316 (2007) 1160.
[31] J.C. Trinidad, C.G. Specht, A. Thalhammer, R. Schoepfer, A.L. Burlingame, Mol. Cell. Proteomics 5 (2006) 914.
[32] A.J. Alpert, G. Mitulović, K. Mechtler, poster presented at 2008 ABRF, Meeting (2008).
[33] S.A. Beausoleil, M. Jedrychowski, D. Schwartz, J.E. Elias, J. Villen, J. Li, M.A. Cohn, L.C. Cantley, S.P. Gygi, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 12130.
[34] C.S. Gan, T. Guo, H. Zhang, S.K. Lim, S.K. Sze, J. Proteome Res. 7 (2008) 4869.
[35] T.E. Thingholm, O.N. Jensen, M.R. Larsen, Methods Mol. Biol. 527 (2009) 67.
[36] M.J. Chalmers, K. Hakansson, R. Johnson, R. Smith, J. Shen, M.R. Emmett, A.G. Marshall, Proteomics 4 (2004) 970.
[37] M.J. Schroeder, D.J. Webb, J. Shabanowitz, A.F. Horwitz, D.F. Hunt, J. Proteome Res. 4 (2005) 1832.
[38] J.A. Dowell, D.C. Frost, J. Zhang, L. Li, Anal. Chem. 80 (2008) 6715.
[39] A.J. Alpert, S.P. Gygi, A.K. Shukla, Poster# MP438, 55th ASMS Conference, June 2007 (2007).