research articles Journal of proteome research

An Automated Platform for Analysis of Phosphoproteomic Datasets: Application to Kidney Collecting Duct Phosphoproteins

Jason D. Hoffert,[†,‡] Guanghui Wang,[‡,§] Trairak Pisitkun,[†] Rong-Fong Shen,[§] and Mark A. Knepper*[,†]

Laboratory of Kidney and Electrolyte Metabolism, Proteomics Core Facility, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892

Received March 2, 2007

Large-scale phosphoproteomic analysis employing liquid chromatography-tandem mass spectrometry (LC−MS/MS) often requires a significant amount of manual manipulation of phosphopeptide datasets in the post-acquisition phase. To assist in this process, we have created software, PhosphoPIC (PhosphoPeptide Identification and Compilation), which can perform a variety of useful functions including automated selection and compilation of phosphopeptide identifications from multiple MS levels, estimation of dataset false discovery rate, and application of appropriate cross-correlation (XCorr) filters. In addition, the output files generated by this program are compatible with downstream phosphorylation site assignment using the Ascore algorithm, as well as phosphopeptide quantification via QUOIL. In this report, we utilized this software to analyze phosphoproteins from short-term vasopressin-treated rat kidney inner medullary collecting duct (IMCD). A total of 925 phosphopeptides representing 173 unique proteins were identified from membrane-enriched fractions of IMCD with a false discovery rate of 1.5%. Of these proteins, 106 were found only in the membrane-enriched fraction of IMCD cells and not in whole IMCD cell lysates. These identifications included a number of well-studied ion and solute transporters including ClC-1, LAT4, MCT2, NBC3, and NHE1, all of which contained novel phosphorylation sites. Using a label-free quantification approach, we identified phosphoproteins that changed in abundance with vasopressin exposure including aquaporin-2 (AQP2), Hnrpa3, IP_3 receptor 3, and pur-beta.

Keywords: phosphoproteomics • neutral loss • target-decoy • LC−MS/MS • collecting duct • IMCD • mass spectrometry • label-free • PhosphoPIC • proteomics

Introduction

Large-scale "shotgun" LC−MS/MS analysis is a powerful method that can be used to identify and quantify phosphopeptides from a complex protein sample.[1−5] This method usually involves analysis of a mixture of phosphopeptides obtained from either whole cell lysates or complex subcellular fractions. With the current availability of high-resolution and high-sensitivity mass spectrometers, MS datasets generated by this particular phosphoproteomic approach have grown progressively larger and more cumbersome. In addition, there are a number of attributes unique to phosphopeptide datasets that create additional challenges in the post-acquisition phase.

One distinctive characteristic of phosphopeptide datasets is that they often include both phosphorylated and nonphosphorylated peptides. Even with the addition of initial chromatography steps such as strong cation exchange and immobilized metal affinity chromatography (IMAC) to enrich for phosphopeptides, a significant number of nonphosphorylated peptides will remain in the sample during MS analysis. Removing nonphosphorylated peptide identifications *in silico* is necessary for both accurate estimation of dataset false discovery rate (FDR) and appropriate application of spectral quality filters (i.e., XCorr, RSp). The presence of a relatively high number of "single peptide" identifications in phosphopeptide datasets heightens the need for appropriate filtering and accurate FDR estimation.

The inclusion of phosphopeptide data from multiple MS levels (i.e., MS^2 and MS^3) also complicates post-acquisition analysis. Higher quality MS^3 spectra are utilized to confirm poorer quality MS^2 identifications,[6] and MS^3 data often contain a significant number of unique phosphopeptides not found at the MS^2 level. However, the presence of the same phosphopeptide in both MS^2 and MS^3 scans can lead to redundant analysis if these files are analyzed separately. To simplify this process, both MS^2 and MS^3 datasets should be analyzed as a single integrated file.

While these data manipulations can be performed manually, the process is both labor-intensive and prone to user error. Programs that automate particular aspects of phosphopro-

* To whom correspondence should be addressed. Mark A. Knepper, M.D. Ph.D., NIH Bldg. 10, Room 6N260, 10 CENTER DR MSC 1603, Bethesda, MD 20892-1603; Phone, (301)496-3064; Fax, (301)402-1443; E-mail, knep@helix.nih.gov.
[†] Laboratory of Kidney and Electrolyte Metabolism.
[‡] These authors contributed equally to this manuscript.
[§] Proteomics Core Facility.

10.1021/pr0701153 Not subject to U.S. Copyright. Publ. 2007 Am. Chem. Soc.

Published on Web 08/07/2007

teomic data analysis are currently available.[2,3,7,8] However, we were especially interested in creating a more general workflow that suited our needs in the analysis of phosphopeptide identifications generated by the SEQUEST search algorithm.

To create a more integrated, user-friendly platform, we developed PhosphoPIC, a Windows-based software suite written in Visual C++ that utilizes SEQUEST search results generated from multiple MS levels to create a single filtered phosphopeptide dataset. This software provides an intuitive linear workflow by including individual programs for most steps of post-acquisition analysis: phosphopeptide selection, FDR estimation, spectral quality filtering, and compilation of MS^2 and MS^3 search results. In this study, PhosphoPIC was successfully used to filter and compile phosphopeptide datasets generated from an LC−MS/MS shotgun analysis of membrane-enriched samples of renal inner medullary collecting duct (IMCD). The software was also useful in generating compatible files for automated phosphorylation site assignment and label-free phosphopeptide quantification.

Materials and Methods

Sample Preparation. IMCD suspensions were prepared from rat kidney inner medullas using the method of Stokes et al.[9] with modifications.[10] After isolation, IMCD suspensions were incubated for 10 min in the absence or presence of 10 nM [deamino-Cys,^{1}D-Arg^8]vasopressin (dDAVP), a V2-receptor-specific analog of the anti-diuretic hormone, arginine vasopressin. After centrifugation at 10 000× g for 30 s, pelleted IMCD tubules were sonicated in ice-cold homogenization buffer: 250 mM sucrose, 10 mM triethanolamine, 1× Complete protease inhibitor cocktail (Roche, Mannheim, Germany), 1× HALT phosphatase inhibitor (Pierce, Rockford, IL), pH adjusted to 7.6. Samples were spun at 1000× g for 10 min to pellet unbroken cells and debris. The supernatant from this step was centrifuged at 200 000× g for 1 h. This "membrane-enriched" pellet (200−600 μg protein) was resuspended in 6 M guanidine-HCl/50 mM ammonium bicarbonate followed by reduction, alkylation, trypsin digestion, and desalting as previously described.[4] In this study, no detergents were used to solubilize membrane protein samples since their presence did not increase the yield of identified phosphopeptides (data not shown). To enrich for phosphopeptides, desalted samples in 0.1% formic acid were subjected to IMAC (Pierce, Phosphopeptide Isolation Kit) using the supplied protocol with slight modifications.[4] Samples were then desalted using C18 Ziptips (Millipore) prior to analysis by mass spectrometry.

LC−MS/MS Analysis. Samples were analyzed on an Agilent 1100 nanoflow LC system (Palo Alto, CA) connected to a Finnigan LTQ FT mass spectrometer (Thermo, San Jose, CA) as previously described.[4] Briefly, the peptide mixture injected by the autosampler was loaded onto a C18 precolumn (Agilent) for desalting. Captured peptides were directed to a PicoFrit reversed-phase analytical column (New Objective, Woburn, MA) and eluted by using a linear gradient of acetonitrile (0−60% in 45 min) in 0.1% formic acid at a flow rate of 250 nL/min. Eluted peptides were ionized and analyzed using the LTQ-FT mass spectrometer equipped with a nanospray ion source. Spectra were recorded in data-dependent acquisition mode and the dynamic exclusion option was enabled. Survey MS scans were acquired by either the LTQ (linear ion trap) or the FT-ICR (Fourier transform-ion cyclotron resonance) detector (with a resolution of 100 000 at m/z of 400), and each survey MS scan was followed by five MS/MS scans using the LTQ. MS^3 scans were triggered if a neutral-loss fragment (loss of 98, 49, or 32.7 from the precursor ion) was observed in the corresponding MS/MS scans.

Target-Decoy Database Search. MS/MS spectra were searched using BioWorks 3.1 (Thermo Scientific) running the SEQUEST algorithm on a concatenated database containing both the forward (target) and reversed complement (decoy) of the most current version of the Rat Refseq Database (National Center for Biotechnology Information). The following parameters were used: 2 missed trypsin cleavage sites permitted, fixed modification of 57.0214 on cysteine, variable modifications of 15.9949 on methionine and 79.9663 on serine, threonine, and tyrosine. For MS^3 spectra, an additional variable modification of −18.0106 on serine and threonine was added to account for dehydration of these two amino acids due to neutral loss of phosphoric acid.

Phosphopeptide Analysis using PhosphoPIC. PhosphoPIC was created with Visual C++ as a set of Microsoft Windows-based programs with output files generated in a flexible, tab-delimited text format. (This software is freely available for noncommercial purposes upon request.) Through the "Phosphopeptide Selection/Filtering" module of PhosphoPIC, all.dta and.out files contained in the SEQUEST search results directory were used to generate a filtered dataset which was constrained by various parameters including the presence of at least one phosphorylated amino acid, a maximum preliminary score rank (RSp) = 1, and a minimum deltaCn = 0. Concomitantly, the PhosphoPIC software automatically adjusted the minimum cross-correlation (XCorr) filter for each individual charge state (+1, +2, +3) separately in order to optimally meet predetermined target false discovery rates (FDRs) based on the number of allowable random matches from the reversed "decoy" database. The actual FDR was calculated as 2R/(F+R) and the adjusted FDR was calculated as R/F, where R is the number of reversed hits and F is the number of forward hits. Setting the maximum RSp value as low as possible allowed PhosphoPIC to choose lower XCorr cutoff values, resulting in a higher number of identifications for a given target FDR. However, filtering for a maximum RSp = 1 excludes possible "true positive" identifications that were "first matches" (highest XCorr value) yet did not have the highest preliminary scores. Thus, there is an obvious tradeoff between coverage and quality when filtering phosphopeptide datasets. The Phosphopeptide Selection/Filtering program was designed for general purpose filtering of peptide identifications at specified target FDRs. Thus, it is also compatible with LC−MS/MS analysis of non-phosphorylated peptides, although not utilized in this capacity in this study.

The end result of this initial filtering process was the creation of two separate spreadsheet files, a multiconsensus report similar to ones obtained through Bioworks, as well as a summary file containing useful information such as XCorr filters applied, actual FDR, adjusted FDR (calculated after removal of random matches), total redundant and nonredundant peptide count for both forward and reversed databases, and number of proteins identified for a given peptide count. MS^2 and MS^3 datasets filtered with a target FDR = 2% were used for all subsequent phosphopeptide analysis. The optimal XCorr cutoff values established for this FDR were 2.02 (+1), 2.92 (+2), and 4.40 (+3) for MS;[2] 1.58 (+1), 2.87 (+2), and 3.23 (+3) for MS.[3] Adjusted FDRs for MS^2 and MS^3 datasets were 0.67 and 0.81%, respectively.

After phosphopeptide selection and filtering, datasets from both MS levels were merged using the "MSn Compilation" module of PhosphoPIC, which generated another pair of

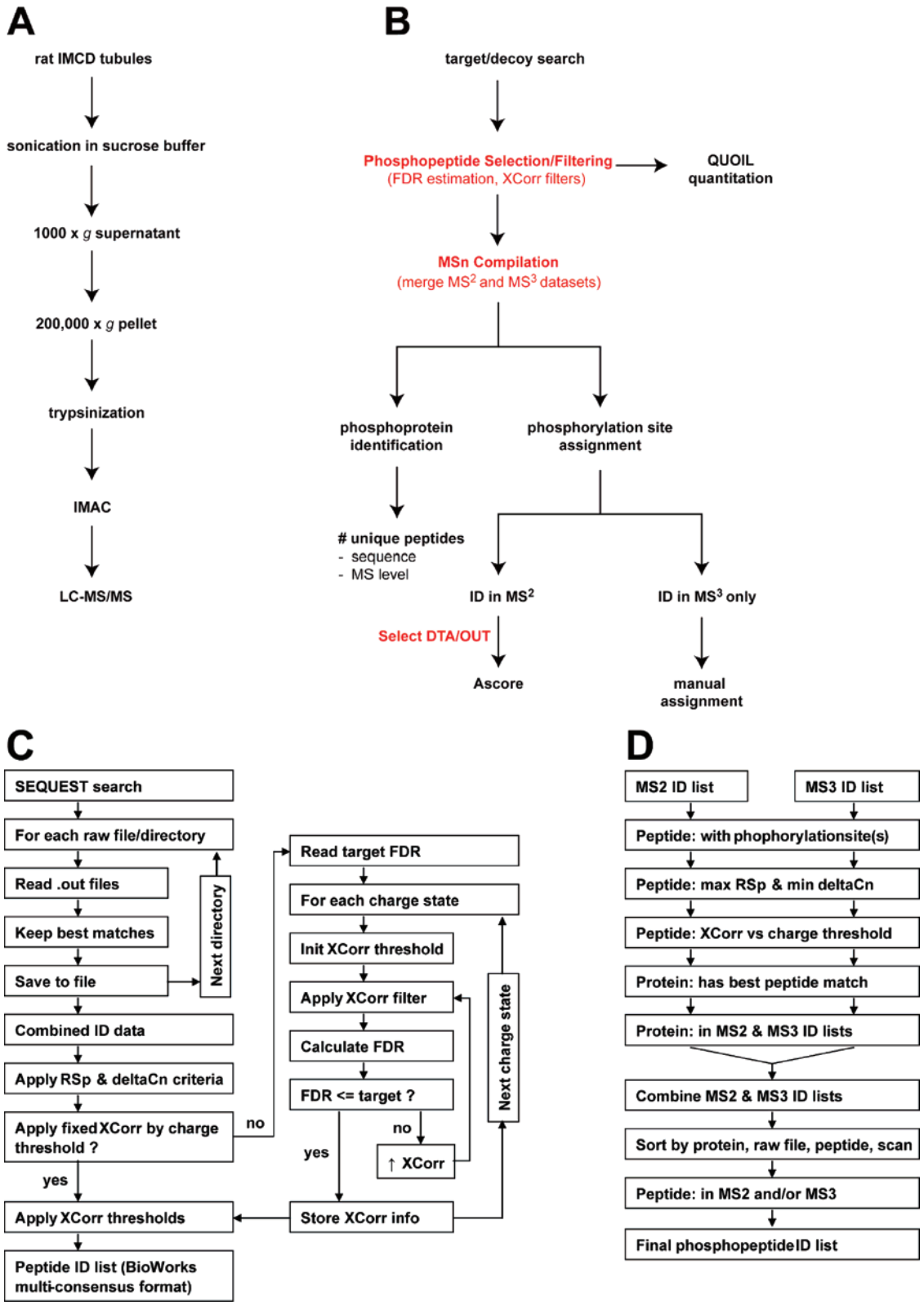
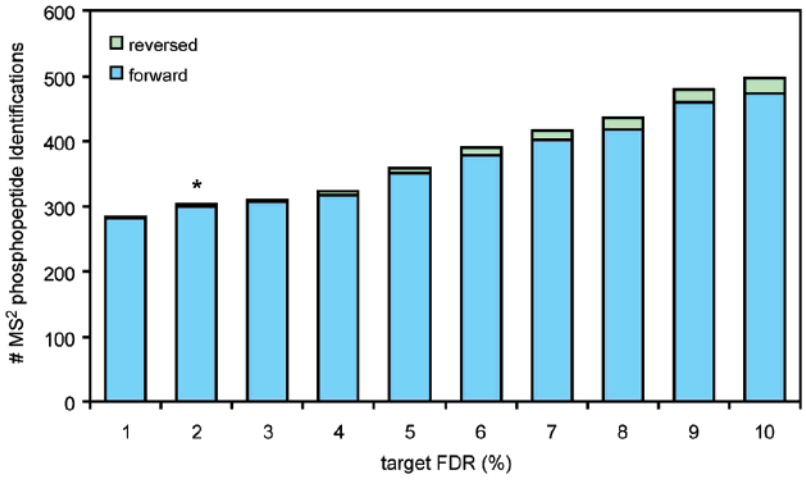


Figure 1. Experimental approach. (A) IMCD cells were enriched from rat kidney inner medullas by enzymatic digestion followed by low-speed centrifugation. Cells were lysed by sonication in sucrose buffer. Differential centrifugation produced a 200 000 g pellet which was digested with trypsin, followed by IMAC to enrich for phosphopeptides. Phosphopeptides were then analyzed by LC−MS/MS. (B) Post-acquisition analysis. The 3 modules that comprise PhosphoPIC are indicated in red font. FDR = false discovery rate; XCorr = cross-correlation filter. (C) Detailed workflow for the "Phosphopeptide Selection/Filtering" module of PhosphoPIC. (D) Detailed workflow for the "MSn Compilation" module of PhosphoPIC.

multiconsensus and summary files. During dataset compilation, optional parameters could be set (e.g., requiring a peptide to be identified in both MS^2 and MS^3 within a certain retention time window, or selecting phosphopeptides identified in MS^3

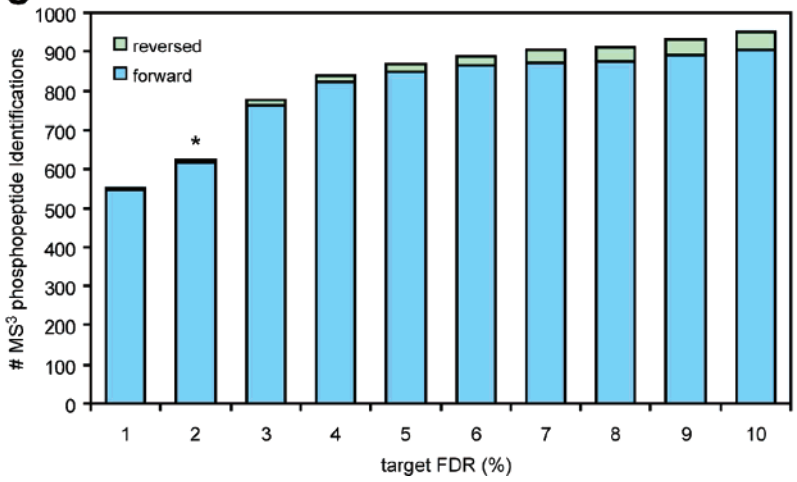
A



B

forward hits	reversed hits	target FDR (%)	actual FDR (%)	adjusted FDR (%)
282	1	1.00	0.71	0.35
300	2	2.00	1.32	0.67
306	3	3.00	1.94	0.98
317	6	4.00	3.72	1.89
351	8	5.00	4.46	2.28
379	11	6.00	5.64	2.90
401	14	7.00	6.75	3.49
418	17	8.00	7.82	4.07
459	21	9.00	8.75	4.58
473	24	10.00	9.66	5.07

C



D

forward hits	reversed hits	target FDR (%)	actual FDR (%)	adjusted FDR (%)
548	2	1.00	0.73	0.36
618	5	2.00	1.61	0.81
764	11	3.00	2.84	1.44
821	16	4.00	3.82	1.95
847	21	5.00	4.84	2.48
864	25	6.00	5.62	2.89
872	31	7.00	6.87	3.56
875	35	8.00	7.69	4.00
890	41	9.00	8.81	4.61
904	45	10.00	9.48	4.98

Figure 2. Estimation of false discovery rates. Both MS^2 (A, B) and MS^3 (C, D) phosphopeptide datasets were filtered using target false discovery rates (FDR) from 1 to 10%. The number of peptides identified from the forward database (blue) as well as from the reversed database (green) increases as the FDR increases owing to the application of less stringent XCorr filters. A target FDR of 2% (*), which corresponded to an actual FDR of 1.32% for MS^2 and 1.61% for MS,^{3} was used for subsequent phosphopeptide identification and site assignment. Adjusted FDR is calculated after removal of hits from the reversed database.

only). These files were used to determine which phosphopeptides were identified at each MS level. Phosphopeptides from MS^2 spectra were selected using PhosphoPIC and the corresponding.dta and.out files for those identifications were analyzed using Ascore (http://ascore.med.harvard.edu/ascore.php), a statistical algorithm that measures the probability of correct phosphorylation site localization based on the presence and intensity of site-determining ions in MS^2 spectra.[1] Phosphorylation sites with an Ascore ≥ 19 (corresponding to >99% certainty) were considered unambiguously assigned. Since the Ascore algorithm is only compatible with MS^2 spectra, phosphorylation sites from phosphopeptides found only at the MS^3 level were assigned by manual examination of the spectra.

Label-Free Quantification. Label-free phosphopeptide quantification was performed using the QUOIL (quantification without isotopic labeling) program[11] which can accept summary files generated by PhosphoPIC for individual sample trials. QUOIL quantifies reconstructed peak ion chromatograms from separate LC−MS/MS runs and has been used successfully for quantification of both non-phosphoryated[11] and phosphorylated peptides.[4] To ensure accurate quantification, MS data were collected using the high-resolution FT-ICR analyzer for precursor ion scanning while the LTQ was used for analysis of fragment ion spectra. The following parameters were used for label-free quantification of MS data: maximum RSp = 2, minimum deltaCn = 0, target FDR = 2%, precursor m/z = theoretical (for MS^2), measured (for MS^3); peptide ratio calculated using peak area; peptide ratio normalization by median at raw file level; protein ratio calculated using median; precursor m/z tolerance = 0.05 Da; peak window = 1 min; minimal S/N = 1; maximal FWHM = 0.3 min; min ratio change: 0.25; peptide set = union, peak RT tolerance between samples = 3 min.

Results and Discussion

General Strategy for Isolation and Analysis of Phosphopeptides. Rat inner medullary collecting duct (IMCD) tubules were enriched from whole inner medulla by a combination of collagenase digestion and differential centrifugation as detailed in the "Materials and Methods" section. Following short-term vasopressin treatment (10 nM dDAVP for 10 min), samples were processed as shown in Figure 1A. Cells were disrupted by

Table 1. Summary of IMCD Phosphopeptide Data Analyzed by PhosphoPIC

Total phosphopeptides	925
Nonredundant phosphopeptides	268
False positives	7
Actual FDR	1.5%
Adjusted FDR	0.76%
Total proteins	173
Proteins in MS^2 only	22
Proteins in MS^3 only	101
Proteins in MS^2 and MS^3	50
Proteins unique to the membrane-enriched fraction of IMCD	106

sonication, followed by differential centrifugation to produce a "membrane-enriched" 200 000 g pellet. This pellet was resuspended in guanidine and digested with trypsin. Phosphopeptides were enriched from total peptide digests by using IMAC prior to LC−MS/MS analysis.

Post-acquisition analysis was performed according to the scheme presented in Figure 1B. Both MS^2 and MS^3 spectra were searched using the target-decoy approach,[1,12] which employed a concatenated database that included both forward and reversed sequences of the latest version of the rat Refseq database. The target-decoy approach is the prevailing method for estimating false discovery rates of MS datasets and is based on the principle that incorrect spectral matches have an equal probability of occurring in either forward ("target") or reversed ("decoy") databases. MS search results were then analyzed using PhosphoPIC. As described more thoroughly in "Materials and Methods", these programs perform 3 main functions: phosphopeptide selection/filtering, MSn compilation, and.dta/.out file selection. The first two functions help to organize, filter, and summarize phosphopeptide identifications, whereas the third function is used to generate the input file for automated phosphorylation site assignment. A more detailed workflow for the Phosphopeptide Selection/Filtering and MSn Compilation modules of PhosphoPIC is given in Figure 1C and D, respectively.

Phosphopeptide Selection/Filtering. To begin the PhosphoPIC analysis, MS datasets were filtered to remove non-phosphorylated peptide identifications. Appropriate XCorr filters were also applied, iteratively, based on user-defined "target FDR" values from 1 to 10%. As shown in Figure 2, as the target FDR was increased, the number of phosphopeptide identifications from both the forward and reversed database increased due to progressively lower XCorr cutoff values. When compared with MS^2 phosphopeptide totals, the MS^3 dataset consistently included a higher number of identifications for all corresponding target FDRs, highlighting the importance of analyzing data from multiple MS levels. The "actual FDR" takes into account an equal number of known and predicted random hits from the reversed and forward databases, respectively. The "adjusted FDR" excludes the random matches from the reversed database and reflects the smallest FDR value that can be estimated by the target-decoy approach. For subsequent analysis, we selected MS^2 and MS^3 phosphopeptide datasets that were filtered to produce a target FDR = 2%, which corresponded to adjusted FDRs < 1%. Single files containing both summary and multi-consensus files are available for MS^2 (Supporting Information Table 1) and MS^3 (Supporting Information Table 2) datasets.

MSn Compilation. Following initial XCorr filtering and FDR estimation, MS^2 and MS^3 datasets were combined into a single

Table 2. Subset of Integral Membrane/Membrane-Associated IMCD Phosphoproteins Identified in This Study[a]

name	accession #	gene	function	family	site[b]
arfaptin 1	NP_068531	Arfip1	Membrane Trafficking Proteins	ah domain	S39*
ATPase, Class VI, type 11B	XP_342229	Atp11b	Transporters and Channels	cation transport ATPase (p-type)	S836*
ATPase, H+ transporting, V1 subunit G, isoform 3	XP_213980	Atp6v1g3	Transporters and Channels	V-ATPase	S234*
ATP-binding cassette sub-family F member 1 (ATP-binding cassette 50)	XP_001056151	Abcf1	Transporters and Channels	ABC transporter	S109
chloride channel 1 (ClC-1)	NP_037279	Clcn1	Transporters and Channels	chloride channel (tc 1.a.11)	T636*, S642*, S645*
claudin-8	NP_001032863	Cldn8	Tight Junction Proteins	claudin	S220*
estrogen receptor-binding fragment-associated gene 9	NP_001020068	Ebag9	Membrane Receptors	coiled coil	S36*
G protein-coupled receptor kinase interactor 1 (GIT1)	NP_114002	Git1	Small GTP Binding Proteins and Related Proteins	arf-gap	S601*
guanine nucleotide binding protein (G protein), gamma 10	XP_577288	Gng10	Heterotrimeric G-proteins	g protein gamma	S49*
membrane-associated guanylate kinase-related (MAGI-3)	NP_620784	Magi3	Signal Proteins	PDZ domain	S829∧, S831∧
PDZ-domain protein scribble	XP_343267	Scrib	Anchor Scaffold Proteins	lap (lrr and pdz)	S506*
secretory carrier membrane protein 3	XP_342280	Scamp3	Membrane Trafficking Proteins	scamp	S78*
solute carrier family 16, member 7 (MCT-2)	NP_058998	Slc16a7	Transporters and Channels	monocarboxylate porter (tc 2.a.1.13)	S488*
solute carrier family 4, member 1 (band 3 anion exchanger, AE1)	NP_036783	Slc4a1	Transporters and Channels	anion exchanger (tc 2.a.31)	S18
solute carrier family 4, member 7 (NBC3)	NP_478118.1	Slc4a7	Transporters and Channels	anion exchanger (tc 2.a.31)	S89*
solute carrier family 43, member 2 (LAT4)	NP_775564	Slc43a2	Transporters and Channels	MFS	S274*
solute carrier family 9, member 1 (NHE1)	NP_036784	Slc9a1	Transporters and Channels	na(+)/h(+) exchanger (tc 2.a.36.5.1)	S697*, S707, S701*, S727*, S730*, S790*
synaptosomal-associated protein 23 (SNAP-23)	NP_073180	Snap23	Membrane Trafficking Proteins	snap-25	S110
syntaxin 7	NP_068641.2	Stx7	Membrane Trafficking Proteins	syntaxin	T79∧
tight junction protein 2 (ZO-2)	NP_446225	Tjp2	Tight Junction Proteins	maguk	S463*
translocase of outer mitochondrial membrane 70	NP_997684	Tomm70a	Translational Proteins	tom70	S94*
vinculin	XP_223781	Vcl	Adhesion Proteins and Collagens	vinculin/alpha-catenin	S626*

[a] Full list is presented in Supporting Information Table 4. [b] The symbol * indicates a novel site, ∧ indicates an ambiguous site.

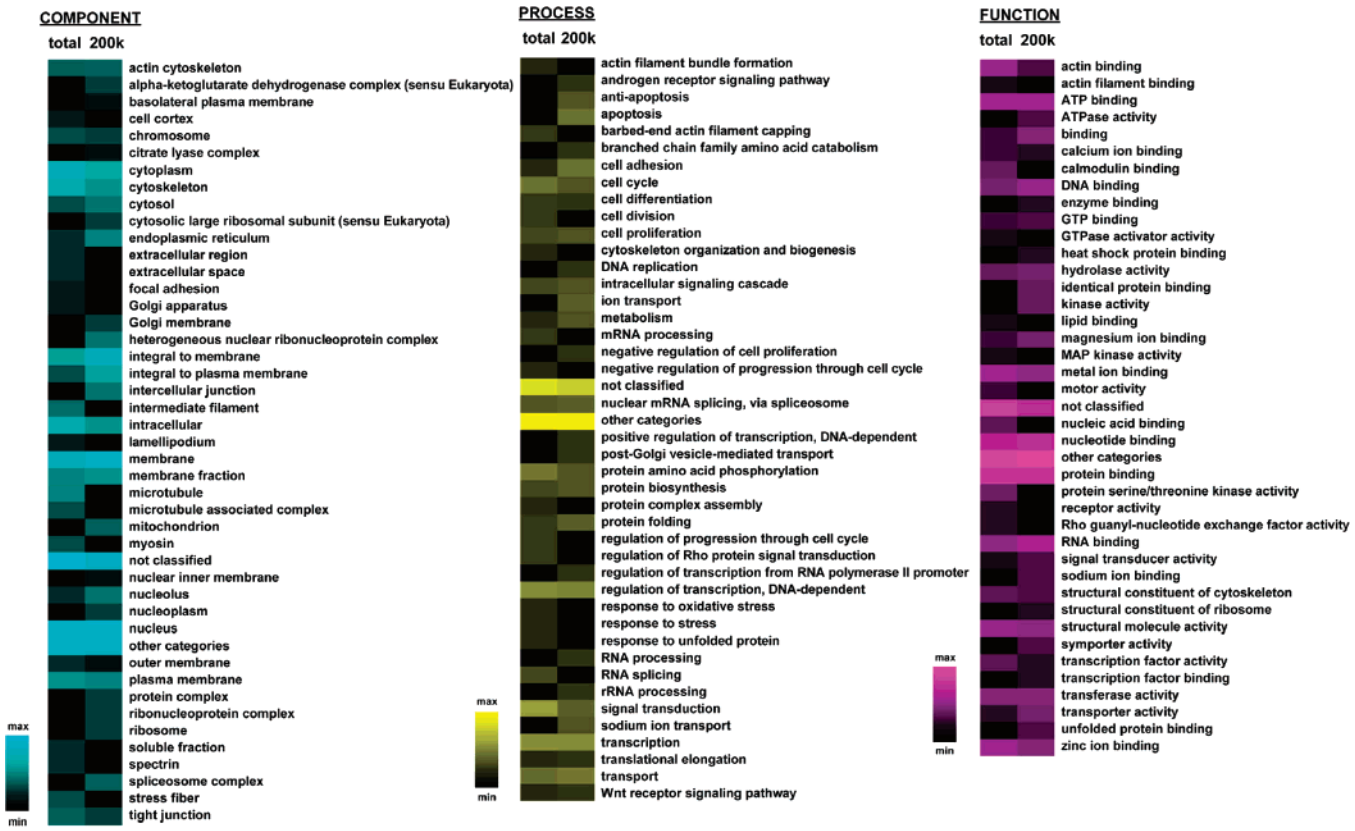


Figure 3. Functional classification of identified IMCD phosphoproteins. Heat maps were constructed from the three major Gene Ontology (GO) categories: cellular component (cyan), biological process (yellow), and molecular function (magenta). The percentage of phosphoproteins associated with each individual GO term has been $\log_2$ transformed and then converted into a 16-bit grayscale intensity value (min = 0, max = 255). A direct comparison of whole cell lysate (total) with the membrane-enriched fraction (200k) indicates an increased number of phosphoproteins in the 200k fraction that have GO terms associated with membrane localization/membrane transport functions.

output file (Supporting Information Table 3). A total of 925 redundant (268 nonredundant) phosphopeptides were identified, including 7 false positive matches (1.5% actual FDR) (Table 1). After removing false positives, 173 unique phosphoproteins remained. The majority of these proteins (126) were identified from single peptide matches, a common characteristic of phosphopeptide datasets. The fact that most phosphoproteins were identified only at the MS^3 level (101) is in contrast to results from a recent large-scale phosphoproteomic analysis of nocodazole-arrested HeLa cell lysates.[1] In that study, Beausoleil et al. reported that only 8.7% of all nonredundant phosphorylation sites were contributed solely by MS^3 spectra. The reason for this discrepancy in the relative contribution of MS^3 spectra to an entire phosphopeptide dataset is not yet known.

Of the 173 phosphoproteins identified from membrane-enriched IMCD samples, 106 of these proteins had not been identified during prior phosphoproteomic analysis of total IMCD cell lysates.[4] These identifications have been added to the online Collecting Duct Phosphoprotein Database (CDPD) (http://dir.nhlbi.nih.gov/papers/lkem/cdpd/).[4] Newly identified integral membrane phosphoproteins included a number of well-known collecting duct ion and solute transport proteins ClC-1, LAT4, MCT2, NBC3, and NHE1 (Table 2). The possibility that some of these new identifications resulted simply from run-to-run LC−MS/MS variability cannot be excluded. A functional classification of all 106 proteins identified from membrane-enriched IMCD is presented as a set of three heat maps based on established Gene Ontology (GO) terms (Figure 3), one for each major GO category (cellular component, biological process, and molecular function). The intensity values correspond to the number of distinct proteins associated with a particular GO term. As expected, GO terms relating to membrane localization (e.g., basolateral plasma membrane, integral to plasma membrane, membrane, membrane fraction) as well as membrane transport (e.g., ion transport, sodium ion transport, transport, symporter activity, transporter activity) were more prominent in the membrane-enriched ("200k") fraction when compared to total IMCD cell lysates. The membrane-enriched fraction also included a large number of nuclear, ER, and mitochondrial proteins, most likely due to the nature of the homogenization and the lack of multiple lower speed centrifugation steps. A summary of all identified phosphoproteins including gene names, accession numbers, and relevant ontology data is included in Supporting Information Table 4.

Phosphorylation Site Assignment. To distinguish the correct site(s) of phosphorylation for each phosphopeptide, automated site assignment was performed on MS^2 data using the Ascore algorithm.[1] The Ascore value for each phosphopeptide is included in Supporting Information Table 3. In all, 59 distinct phosphorylation sites were unambiguously assigned (Ascore $\geq$ 19). For the most part, these sites were identical to those found in corresponding MS^3 spectra of the same peptide. Manual spectral validation was performed on phosphopeptides found only at the MS^3 level. Of the 140 nonredundant phos-

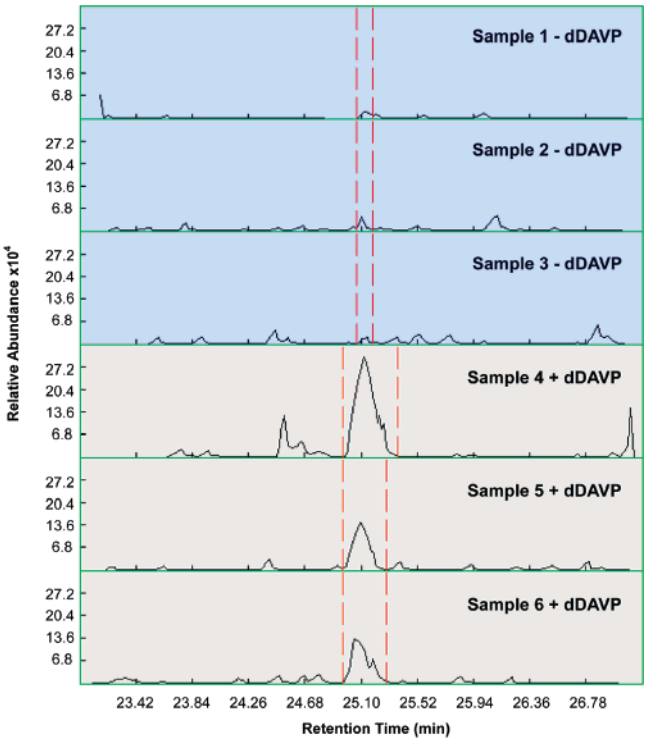


Figure 4. Label-free quantification using QUOIL software. Phosphopeptides from IMCD samples incubated in the absence or presence of 10 nM dDAVP for 10 min were selected using PhosphoPIC software and then quantified using a label-free approach based on normalized peak area ratios obtained from reconstructed ion chromatograms. Shown are the extracted ion chromatograms for a triply phosphorylated AQP2 peptide (RQpSVELHpSPQpSLPR). The dashed red lines indicate the beginning and the end of the peaks that were quantified. The relative abundance of this AQP2 phosphopeptide was increased an average of 20.7-fold in samples that were incubated with dDAVP ($n = 3$).

phorylation sites among the 106 phosphoproteins unique to the membrane-enriched fraction of rat IMCD, 41 had been previously identified and 76 were novel according to the online phosphoprotein database PhosphoSite (www.phosphosite.org). (PhosphoSite includes phosphorylation site information mostly for human and mouse proteins. If a site corresponding to the rat sequence was identified in mouse and/or human, then this site was considered "previously identified" even if it had not been identified in rat.) The remaining 23 sites could not be definitively assigned, although the corresponding phosphopeptides were still useful for phosphoprotein identification. For MS^2 spectra, Ascore values < 19 were considered ambiguous. For MS^3 spectra, an ambiguous assignment resulted from unsuccessful manual inspection. When residues could not be definitively assigned, lack of assignment was usually due to a lack of distinguishing ion peaks surrounding the potential residue(s) or to generally poor spectral quality. A complete list of all successfully assigned and ambiguous phosphorylation sites is included in Supporting Information Table 4.

Phosphopeptide Quantification. To identify phosphopeptides that changed in abundance due to vasopressin treatment, label-free quantification was performed on IMCD samples incubated with vehicle control ($n = 3$) and IMCD samples incubated with 10 nM dDAVP for 10 min ($n = 3$). Initially, PhosphoPIC was used to generate a separate MS^2 multiconsensus file for each sample. These files were then submitted to QUOIL to allow reconstruction of the corresponding parent ion chromatograms. The entire process was then repeated for MS^3 data. While the majority of phosphopeptides did not change in abundance in the presence of vasopressin, a small number of phosphopeptides showed significant changes including peptides for heterogeneous nuclear ribonucleoprotein A3 (decreased 4.0-fold), IP_3 receptor 3 (decreased 5.8-fold), purbeta (decreased 1.5-fold), and aquaporin-2 (increased 20.7-fold) (Supporting Information Table 5). Of particular relevance to our studies is the decrease in abundance of phosphorylated IP_3 receptor 3. Although the sites of phosphorylation could not be unambiguously assigned, the identified peptide includes Ser-1832, a known target of protein kinase A (PKA),[13,14] which could potentially play a role in vasopressin signaling by modulating the release of intracellular Ca^{+2} from IP_3-sensitive stores.

Another interesting phosphoprotein that was quantified in this study was aquaporin-2 (AQP2), the vasopressin-sensitive water channel.[15,16] Quantification of the AQP2 phosphopeptide ion chromatograms is shown in Figure 4. This particular AQP2 peptide was identified from an MS^3 spectrum and contained three unambiguously assigned phosphorylation sites: Ser-256, Ser-261, and Ser-264. A previous phosphoproteomic study by our group included MS-based quantification of AQP2 at Ser-256 and Ser-261;[4] however, none of the phosphopeptides in that study included phosphorylation at Ser-264. The dramatic increase in abundance of this phosphopeptide in vasopressin-treated samples is consistent with increased phosphorylation of AQP2 at Ser-256 in response to vasopressin.[4,17,18] Although QUOIL has been routinely utilized to reconstruct parent ion chromatograms based on MS^2-level peptide identifications,[4,11] these current results demonstrate the ability of QUOIL to measure relative abundance ratios of phosphopeptides identified from MS^3 spectra as well.

Conclusions

In this paper, we have introduced a new software application, PhosphoPIC, for analysis of phosphopeptides from large-scale LC−MS/MS datasets and have utilized this software to successfully analyze phosphopeptides from a membrane-enriched fraction of rat IMCD. This software was intuitive to use and powerful in its ability to organize, filter, and compile large phosphopeptide datasets from multiple MS levels. A major advantage of using PhosphoPIC is its ability to interface with other programs for analyzing phosphoproteomic data such as QUOIL and Ascore, thus creating an automated platform that encompasses most aspects of phosphopeptide analysis.

Acknowledgment. We thank Dr. Sean Beausoleil for assistance with Ascore analysis. This work was supported by the National Heart, Lung, and Blood Institute intramural budget ZO1-HL001285.

Supporting Information Available: MS^2 summary and multiconsensus files (Supporting Information Table 1); MS^3 summary and multiconsensus files (Supporting Information Table 2); combined MS^2 and MS^3 datasets with Ascore results (Supporting Information Table 3); summary of newly identified IMCD phosphoproteins including gene names, accession numbers, site(s) of phosphorylation, and relevant GO terms (Supporting Information Table 4); relevant statistics for phosphoproteins that changed in abundance with short-term vasopressin treatment (Supporting Information Table 5); instructional files

for using PhosphoPIC (readme_phosphoPIC.txt) and QUOIL (readme_QUOIL.txt). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24*(10), 1285−1292.

(2) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*(3), 635−648.

(3) Lu, B.; Ruse, C.; Xu, T.; Park, S. K.; Iii, J. Y. Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal. Chem.* **2007**, *79*(4), 1301−1310.

(4) Hoffert, J. D.; Pisitkun, T.; Wang, G.; Shen, R. F.; Knepper, M. A. Quantitative phosphoproteomics of vasopressin-sensitive renal cells: regulation of aquaporin-2 phosphorylation at two sites. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*(18), 7159−7164.

(5) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **2005**, *4*(3), 310−327.

(6) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*(37), 13417−13422.

(7) Tao, W. A.; Wollscheid, B.; O'Brien, R.; Eng, J. K.; Li, X. J.; Bodenmiller, B.; Watts, J. D.; Hood, L.; Aebersold, R. Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat. Methods* **2005**, *2*(8), 591−598.

(8) Tabb, D. L.; McDonald, W. H.; Yates, J. R., III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1*(1), 21−26.

(9) Stokes, J. B.; Grupp, C.; Kinne, R. K. Purification of rat papillary collecting duct cells: functional and metabolic assessment. *Am. J. Physiol.* **1987**, *253*(2 Pt 2), F251−F262.

(10) Pisitkun, T.; Bieniek, J.; Tchapyjnikov, D.; Wang, G.; Wu, W. W.; Shen, R. F.; Knepper, M. A. High-throughput identification of IMCD proteins using LC−MS/MS. *Physiol. Genomics* **2006**, *25*(2), 263−276.

(11) Wang, G.; Wu, W. W.; Zeng, W.; Chou, C. L.; Shen, R. F. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J. Proteome Res.* **2006**, *5*(5), 1214−1223.

(12) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat. Biotechnol.* **2004**, *22*(2), 214−219.

(13) Soulsby, M. D.; Wojcikiewicz, R. J. Calcium mobilization via type III inositol 1,4,5-trisphosphate receptors is not altered by PKA-mediated phosphorylation of serines 916, 934, and 1832. *Cell Calcium* **2007**, *42*, 261−270.

(14) Soulsby, M. D.; Wojcikiewicz, R. J. The type III inositol 1,4,5-trisphosphate receptor is phosphorylated by cAMP-dependent protein kinase at three sites. *Biochem. J.* **2005**, *392*(Pt 3), 493−497.

(15) Nielsen, S.; Chou, C. L.; Marples, D.; Christensen, E. I.; Kishore, B. K.; Knepper, M. A. Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membrane. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*(4), 1013−1017.

(16) Nielsen, S.; DiGiovanni, S. R.; Christensen, E. I.; Knepper, M. A.; Harris, H. W. Cellular and subcellular immunolocalization of vasopressin-regulated water channel in rat kidney. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*(24), 11663−11667.

(17) Kuwahara, M.; Fushimi, K.; Terada, Y.; Bai, L.; Marumo, F.; Sasaki, S. cAMP-dependent phosphorylation stimulates water permeability of aquaporin-collecting duct water channel protein expressed in Xenopus oocytes. *J. Biol. Chem.* **1995**, *270*(18), 10384−10387.

(18) Christensen, B. M.; Zelenina, M.; Aperia, A.; Nielsen, S. Localization and regulation of PKA-phosphorylated AQP2 in response to V(2)-receptor agonist/antagonist treatment. *Am. J. Physiol. Renal Physiol.* **2000**, *278*(1), F29−F42.

PR0701153