

Peptide Orientation Affects Selectivity in Ion-Exchange Chromatography

Andrew J. Alpert,^{*,†} Konstantinos Petritis,^{‡,¶} Lars Kangas,[‡] Richard D. Smith,[‡] Karl Mechtler,[§] Goran Mitulović,^{§,#} Shabaz Mohammed,^{⊥,||} and Albert J. R. Heck^{⊥,||}

PolyLC Inc., 9151 Rumsey Road, Ste. 180, Columbia, Maryland 21045, Fundamental and Computational Sciences Directorate, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352, Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria, and Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, and Netherlands Proteomics Centre, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Here we demonstrate that separation of proteolytic peptides, having the same net charge and one basic residue, is affected by their specific orientation toward the stationary phase in ion-exchange chromatography. In electrostatic repulsion–hydrophilic interaction chromatography (ERLIC) with an anion-exchange material, the C-terminus of the peptides is, on average, oriented toward the stationary phase. In cation exchange, the average peptide orientation is the opposite. Data with synthetic peptides, serving as orientation probes, indicate that in tryptic/Lys-C peptides the C-terminal carboxyl group appears to be in a zwitterionic bond with the side chain of the C-terminal Lys/Arg residue. In effect, the side chain is then less basic than the N-terminus, accounting for the specific orientation of tryptic and Lys-C peptides. Analyses of larger sets of peptides, generated from lysates by either Lys-N, Lys-C, or trypsin, reveal that specific peptide orientation affects the ability of charged side chains, such as phosphate residues, to influence retention. Phosphorylated residues that are remote in the sequence from the binding site affect retention less than those that are closer. When a peptide contains multiple charged sites, then orientation is observed to be less rigid and retention tends to be governed by the peptide's net charge rather than its sequence. These general observations could be of value in confirming a peptide's identification and, in particular, phosphosite assignments in proteomics analyses. More generally, orientation accounts for the ability of chromatography to separate peptides of the same composition but different sequence.

Most peptide separations in the past 30 years have involved reversed-phase chromatography (RPC). A number of models have

been developed for prediction of elution of peptides in this mode. For the most part, they involve calculations of the degree to which each amino acid would contribute to retention, its position in the peptide notwithstanding.^{1,2} Some papers have also taken into account a residue's position in the peptide.^{3–5} In the past 10 years, cation exchange of peptides has developed into a significant complement to RPC in multidimensional separation of peptides. It fractionates tryptic peptides into subsets of a size and reduced complexity more manageable for subsequent RPC-MS analysis in proteomics.^{6–8}

A few general rules have been put forth to link a peptide's net charge with its elution time in cation-exchange chromatography (SCX),^{9–12} mainly to eliminate tentative identifications that were inconsistent with the chromatographic behavior. Otherwise, no models have been developed that successfully linked peptides' composition with their behavior in ion-exchange chromatography (IEX). While developing the use of the new ERLIC (electrostatic repulsion–hydrophilic interaction chromatography) mode of chromatography¹³ for the isolation and separation of tryptic phosphopeptides,^{14–16} we observed unexpected selectivity for sequence

- (1) Meek, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1632–1636.
- (2) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Paša-Tolić, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y.; Zhao, R.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 1039–1038.
- (3) Houghten, R. A.; DeGraw, S. T. *J. Chromatogr.* **1987**, *386*, 223–228.
- (4) Petritis, K.; Kangas, L. J.; Yan, B.; Strittmatter, E. F.; Monroe, M.; Qian, W. J.; Adkins, J. N.; Moore, R.; Xu, Y.; Lipton, M. S.; Camp, D. G., II; Smith, R. D. *Anal. Chem.* **2006**, *78*, 5039–5039.
- (5) Tripet, B.; Cepeniene, D.; Kovacs, J. M.; Mant, C. T.; Krokhn, O. V.; Hodges, R. S. *J. Chromatogr. A* **2007**, *1141*, 212–225.
- (6) Sandra, K.; Moshir, M.; D'hondt, F.; Tuytten, R.; Verleysen, K.; Kas, K.; François, I.; Sandra, P. *J. Chromatogr. B* **2009**, *877*, 1019–1039.
- (7) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676–682.
- (8) Motoyama, A.; Yates, J. R., III. *Anal. Chem.* **2008**, *80*, 7187–7193.
- (9) Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.; Goehle, G. R.; Knight, R. D.; Ahn, N. G. *Anal. Chem.* **2004**, *76*, 3556–3568.
- (10) Yen, C.-Y.; Russell, S.; Mendoza, A. M.; Meyer-Arendt, K.; Sun, S.; Cios, K. J.; Ahn, N. G.; Resing, K. A. *Anal. Chem.* **2006**, *78*, 1071–1084.
- (11) Ballif, B. A.; Villén, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Mol. Cell. Proteomics* **2004**, *3*, 1093–1101.
- (12) Trinidad, J. C.; Specht, C. G.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L. *Mol. Cell. Proteomics* **2006**, *5*, 914–922.
- (13) Alpert, A. J. *Anal. Chem.* **2008**, *80*, 62–76.
- (14) Alpert, A. J.; Gygi, S. P.; Shukla, A. K. Desalting Phosphopeptides by Solid-Phase Extraction, Poster# MP438, *55th ASMS Conference*, June, **2007**.

* To whom correspondence should be addressed. E-mail: aalpert@polylc.com.

† PolyLC Inc.

‡ Pacific Northwest National Laboratory.

§ Research Institute of Molecular Pathology.

⊥ Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University.

|| Netherlands Proteomics Centre, Utrecht University.

Present address: Translational Genomic Research Institute, 445 N. Fifth Street, Phoenix, AZ 85004.

Present address: Institute of Cancer Research, Department of Medicine I, Medical University of Vienna, Borschkegasse 8a/A-1090 Vienna, Austria.

variants. The elution trend suggested that peptides can migrate through an IEX column in a well-defined orientation. This study is an attempt to verify that hypothesis. The initial investigation was performed with synthetic peptides with systematically varied sequence. Subsequently, digests of cellular lysates were studied in SCX with numbers of peptides large enough to afford statistically meaningful results. Digests generated with trypsin, Lys-C, or Lys-N were used, both because of the widespread use of such digests and because, for the most part, they contain peptides with charged residues of well-defined number and position. Our combined results reveal that the orientation of peptides is a significant factor in IEX separations.

MATERIALS AND METHODS

ERLIC of Synthetic Peptides: The following peptides were synthesized as described¹⁷ using standard Fmoc solid-phase chemistry on an ABI 433A peptide synthesizer (Applied Biosystems, Foster City, CA): (a) the peptide GGAAGLYG(p)LGK; (b) the set of peptides with the sequence WWGSGPSGSGGGGK, with phosphate groups on 0–2 Ser residues; (c) the sequence variant peptides NAAAAAAWK, AAANAAWK, AAAAAWNK, and their amidated analogues. Peptides SLYSSSPGGAYVTR (Vimentin(51–64)), SLYSS(p)PGGAYVTR, SVNFLSLTPNEIK (MAP 1B(1271–1282)), and SVNFLSLT(p)PNEIK were a gift of Ken Jackson (Molecular Biology-Proteomics Facility, University of Oklahoma Health Sciences Center). ERLIC was performed with a PolyWAX LP column from PolyLC Inc. (Columbia, MD), 5 μm , 300 \AA ; length, 10 or 20 cm.

Artificial Neural Network (ANN) Study of Tryptic Digests: To establish the database used to train and test the ANN, peptide identifications from 1650 LC-MS/MS analyses were filtered according to previously described criteria.⁴ Of the filtered peptides (~300 000), only those that were observed three or more times (~185 000) in different LC-MS/MS analyses were placed in the database. These identifications were obtained from 19 organisms, including bacteria, viruses, mammalian cell lines, tissue homogenates, and body fluids. SCX was performed with a PolySULFOETHYL A column (PolyLC Inc.), 200 \times 2.1 mm, 5 μm , 300 \AA . SCX mobile phase A was 10 mM ammonium formate, pH 3.0, containing 25% acetonitrile (ACN), while mobile phase B was the same with 500 mM ammonium formate. The flow rate was 0.2 mL/min. The gradient was 100% A for 10', then 0–50% B over 40', then 50–100% B over 10', then 100% B for 10'. Detection was monitored at 280 nm, and 25 fractions were collected at uniform intervals during the 70' run. The ANN architecture used in this study had 1053 input nodes, 5 hidden nodes, and 1 output node. The input vector encoded 50 residue positions, each represented as a 21-dimensional bit (sub)vector that corresponded to 20 different unmodified proteogenic amino acids and alkylated cysteines. For a peptide with “*n*” amino acid residues, residue position 1 (i.e., the first 21 bits) encoded the first residue, while residue position 50 (i.e., the last 21 bits) encoded the *n*th residue.

Each succeeding amino acid residue was encoded alternately at positions 2, 49, 3, 48, *etc.* until all residues in the peptide were accounted for. As most peptides are shorter than 50 residues, many of the sequence vector portions did not encode any residues. The input vector also included information pertaining to peptide length, the charge of the peptide in the gas phase, and the peptide net charge in solution at pH 3. The gas-phase charge was computed as the average of the multiple observations of a peptide, provided it was observed in more than one analysis. For example, if a peptide was observed four times with a 2+ charge and four times with a 3+ charge, then the “observed” charge was 2.5. Only peptides that were observed three or more times were allowed in the training database to afford better mean values. Nevertheless, the calculated mean charge states were skewed toward the lower charge states since MS² spectra with lower charge states are more successfully identified in database searches than those with higher charge states. Peptide length, gas-phase charge, and net charge factors were further normalized to one by linear regression so that all input magnitudes were consistent. The total input vector length was thus 1053 inputs: 1050 (50 positions \times 21 bits) for encoding peptide sequences and 3 for encoding the other peptide descriptors (length, gas-phase charge, net charge). For the output node, a normalized value of the SCX retention fraction was used. Causal index (CI) values were used to calculate the contribution of each input to the retention in SCX. Examining the CI for the output as a function of the inputs reveals the direction (positive or negative) and the relative magnitude of the relationship of the inputs on the output. The peptide database was generated from analyses performed using several mass spectrometers, including LCQ Duo, LCQ Deca, LCQ XP, and LTQ (ThermoFinnigan, San Jose, CA) ion trap mass spectrometers. The ANN software, NeuroWindows version 4.5 (Ward Systems Group, USA), utilized a standard back-propagation algorithm on an Intel Xeon workstation.

Lys-N, Lys-C, and Trypsin Digest SCX Experiments: Full experimental details have been described previously.¹⁸ Briefly, HEK 293T (human embryonic kidney) cell lysis was performed, and 1 mg portions were digested with Lys-N (Seikagaku Corp., Tokyo, Japan), Lys-C (Roche Diagnostics, Ingelheim, Germany), or trypsin (Roche Diagnostics). (The list of peptides in the Lys-N, Lys-C, and tryptic digests of the HEK 293T cell lysate is available free of charge via the Internet at <http://pubs.acs.org/doi/suppl/10.1021/ac9004309>.) Peptides from each digest corresponding to 1 mg of protein material were trapped and desalted on C-18 cartridges and then fractionated by SCX as described. A PolySULFOETHYL A column was used, the same as above but with 200 \AA pores. The difference in pore diameter is not a concern here because steric hindrance of peptide diffusion or orientation would not be encountered with either diameter. Fractions were collected in 1 min intervals for 40 min. After evaporation of the solvents, fractionated peptides were resuspended in 60 μL of 10% formic acid. Twenty microliters of each fraction was then analyzed by reversed-phase LC-MS/MS.

(15) Alpert, A.; Mitulović, G.; Mechtler, K. Isolation of Tryptic Phosphopeptides by ERLIC (Electrostatic Repulsion-Hydrophilic Interaction Chromatography), Poster# P-2412-W, *HPLC 2008 Conference*, May 2008.

(16) Gan, C. S.; Guo, T.; Zhang, H.; Lim, S. K.; Sze, S. K. *J. Proteome Res.* **2008**, *7*, 4869–4877.

(17) Mazanek, M.; Mitulović, G.; Herzog, F.; Stingl, C.; Hutchins, J. R. A.; Peters, J.-M.; Mechtler, K. *Nat. Protoc.* **2006**, *1*, 1059–1069.

(18) Gauci, S.; Helbig, A. O.; Slijper, M.; Krijgsveld, J.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2009**, *81*, 4493–4501.

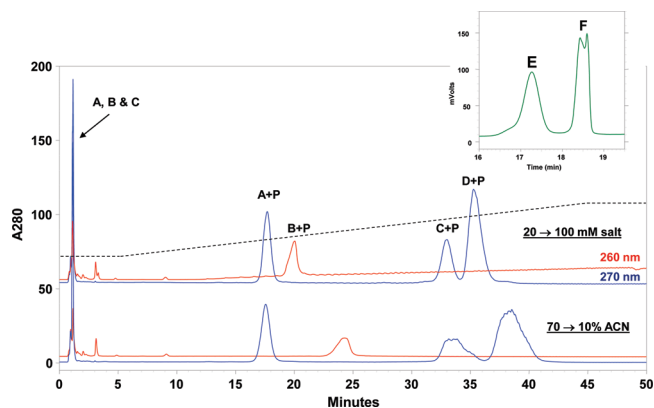


Figure 1. ERLIC of ideal tryptic monophosphopeptide standards; comparison of elution with an increasing salt versus decreasing organic gradient. Column: PolyWAX LP, 100×4.6 mm; $5 \mu\text{m}$, 300 \AA . Flow rate = 1 mL/min . Detection: absorbance; wavelength as noted. Gradient: $0-5'$, 0% B; $5-45'$, $0-100\%$ B; $45-50'$, 100% B. Mobile phase A: 20 mM ammonium formate, pH 2.2, with 70% ACN. Mobile phase B: [top] 100 mM ammonium formate, pH 2.2, with 64% ACN; [bottom] 20 mM ammonium formate, pH 2.2, with 10% ACN. Key: (A) SLYSSSPGGAYVTR (Vimentin(51-64)); (B) SVNFLSLTPNEIK (MAP 1B(1271-1281)); (C) WWGSGPSGSGGGGGK; (A+P) SLYSS(pS)PGGAYVTR; (B+P) SVNFLSL(pT)PNEIK; (C+P) WWGSGPSGSGG(pS)GGGK; (D+P) GGAAGLG(pY)LGK. (Inset) Mobile phase A: 20 mM sodium methylphosphonate, pH 2.0, with 70% ACN. Mobile phase B: 200 mM triethylamine phosphate, pH 2.0, with 60% ACN. Gradient: $0-5'$, 0% B; $5-25'$, $0-100\%$ B. Detection: A220. Column and flow rate: as above. Key: (E) WWGSGP(pS)GSGG(pS)GGGK; (F) WWGSGPSG(pS)GG(pS)GGGK. The double peak for peptide F probably reflects the presence of different conformational isomers around the proline residue.

RESULTS

Sequence Effects in IEX. Recently, a new mode of chromatography, called ERLIC (electrostatic repulsion–hydrophilic interaction chromatography), has been introduced for general-purpose separation of charged solutes.¹³ ERLIC of peptides is performed with an anion-exchange column operated at a pH low enough to uncharge Asp and Glu residues and confer a net positive charge on peptides. While the functional groups of the stationary phase repel the peptides electrostatically, this is

balanced by an appropriate level of hydrophilic interaction. It was assumed that tryptic peptides would migrate in this mode with the middle residues oriented toward the stationary phase (=“down”), reflecting the repulsion by the stationary phase of the positively charged N-terminus at one end and the basic side chain of the Lys or Arg residue at the C-terminus. Phosphate groups retain negative charge under ERLIC conditions, permitting the selective enrichment of phosphopeptides from complex tryptic digests.^{13–16} During a study of the behavior of phosphopeptides with tryptic sequences in the ERLIC mode, it was observed that a random set of such peptides eluted in order of closer proximity of the phosphate group to the C-terminus (Figure 1). The same observation was made regarding the relative elution times of phosphorylation positional variants of a synthetic diphosphopeptide (Figure 1, inset). This trend suggests that, in ERLIC, tryptic peptides tend to be oriented not with the middle of the peptide down but with the C-terminus down. The closer the phosphate group is to the C-terminus, then the closer it is to the positively charged stationary phase and the more it promotes retention (Figure 2, left).

In cation exchange (e.g., SCX) of peptides, the opposite orientation may pertain, with the N-terminus facing the stationary phase and the C-terminus remote from it. Charged residues near the N-terminus would affect retention in SCX more than charged residues near the C-terminus, as portrayed in Figure 2 (right) with phosphate groups. The implication is that a rigid orientation of a peptide in chromatography confers sensitivity to its sequence as well as to the total content of charged residues. If this is true, then peptide orientation is an important factor that must be taken into account in any model of peptide retention in IEX.

Proof of Concept by Using Synthetic Peptides as Orientation Probes in ERLIC. The peptide orientation phenomenon was first confirmed with a series of peptide standards designed to serve as orientation probes. The first set had the sequence NAAAAAWK; AAANAAAWK; AAAAAWLNK. Asparagine is a very polar amino acid, and since ERLIC is a variant of hydrophilic interaction chromatography (HILIC), then addition of an Asn residue will increase retention. Our reasoning was that the peptide that was oriented in a manner that permitted the Asn residue to interact

ORIENTATION OF TRYPTIC PHOSHOPEPTIDES

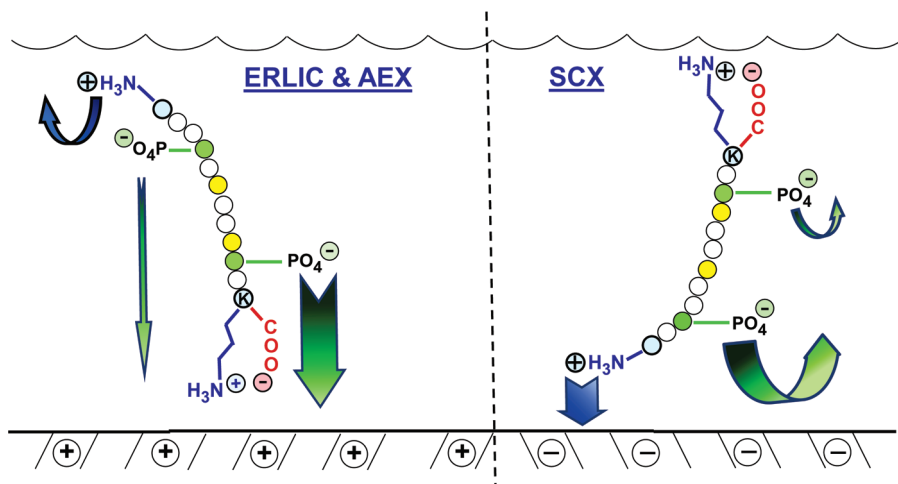


Figure 2. Orientation of tryptic phosphopeptides in ERLIC with an anion-exchange material (left) and in SCX (right).

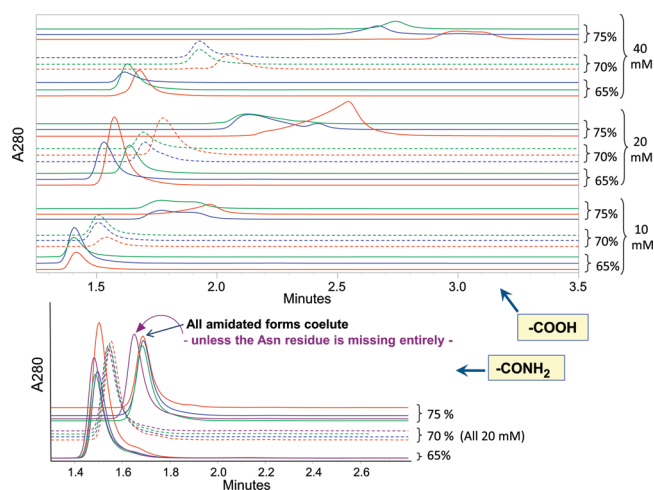


Figure 3. Probes of tryptic peptide orientation in ERLIC. Column: PolyWAX LP, 200 × 4.6 mm; 5 μm, 300 Å. Flow rate = 1 mL/min. Detection: 280 nm. Mobile phase (isocratic): 10, 20, or 40 mM ammonium formate, pH 2.2; %ACN as indicated. Key: Top = AAAAAW \mathbf{N} K (red); AA \mathbf{N} AAAWK (green); NAAAAAWK (blue); Bottom = AAAAAW \mathbf{N} K-NH₂ (red); AA \mathbf{N} AAAWK-NH₂ (green); NAAAAAWK-NH₂ (blue); AAAAAWK-NH₂ (purple).

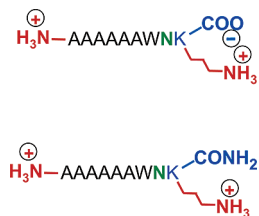


Figure 4. Charge distribution in a regular and an amidated tryptic peptide.

most strongly with the stationary phase would be the last peptide to elute in the ERLIC mode. In addition, the peptides had a C-terminal Lys residue, making it tryptic-like, as well as a Trp residue to permit absorbance detection at longer wavelengths. Figure 3 (top) shows that under various combinations of salt concentration and % ACN, the peptide with the Asn residue closest to the C-terminus nearly always eluted last. These data indicate that these peptides are oriented with the C-terminus toward the stationary phase, with an Asn residue close to the C-terminus affecting retention significantly and an Asn residue in the middle or at the N-terminus having less effect.

We hypothesized that the C-terminus of a tryptic peptide can form a zwitterion with the C-terminal Lys or Arg residue (Figure 4, top). This would leave the side chain of the Lys/Arg residue with less of a net positive charge than would be true of the N-terminus. The C-terminal end of the peptide would then be oriented toward the ERLIC stationary phase not because it was attracted to it but because it experienced less electrostatic repulsion than did the N-terminus. This hypothesis was tested with a second set of peptides, identical to the first set but with the C-termini amidated so as to prevent any electrostatic interaction with the side chain of the Lys/Arg residue (Figure 4, bottom). The Asn residue still promoted retention because its deletion led to earlier elution (Figure 3, bottom). However, all three peptides with the Asn residue coeluted, indicating that any preference in orientation had been abolished.

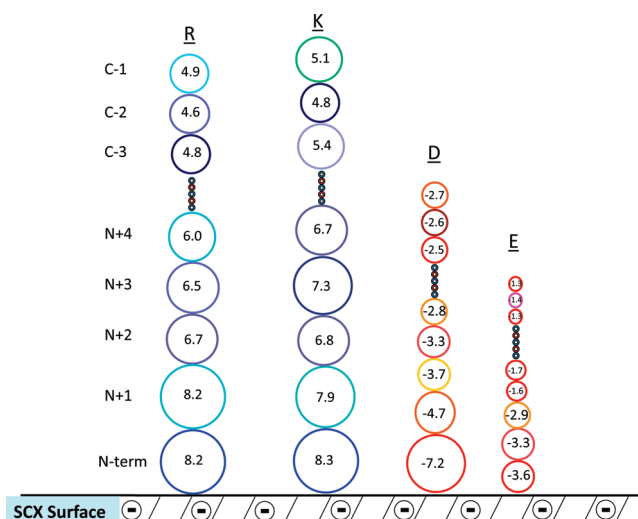


Figure 5. Contribution of a charged residue to retention in SCX as a function of position in a tryptic peptide. Calculations are per ANN comparison of retention times versus sequence (see Materials and Methods for conditions).

These data all support the hypothesis that the basic side chain at the C-terminus of a tryptic peptide forms a zwitterionic bond with the C-terminus itself and is therefore less basic than the N-terminus.

Degree of Influence of a Charged Residue at a Terminus in SCX: ANN Analysis. If the N-terminus of a tryptic peptide is more basic than the functional group of the Lys/Arg residue at the C-terminus, then one would expect the orientation in cation exchange to be the opposite of that in anion exchange or ERLIC, with the N-terminus down. This hypothesis was examined with a set of ~185 000 tryptic peptides separated by SCX in ~1650 LC-MS/MS analyses. The resulting list was mined by an artificial neural network (ANN) in a search for any relationship between basic and acidic residue locations within a peptide and its retention. The schematic in Figure 5 portrays the degree to which charged residues at various positions affected retention. An Asp residue at the N-terminus significantly decreases retention, presumably by forming a zwitterionic bond with the N-terminus and impairing its interaction with the stationary phase. This effect decays rapidly with Asp residues located at increasingly distant positions. An Asp residue located next to the C-terminal Lys or Arg residue affected retention no more than an Asp residue at a remote location in the middle of the peptide. This implies that the peptides are in fact highly oriented with the N-terminus, not the C-terminus, in contact with the stationary phase. The effect with a Glu residue was somewhat similar although less pronounced. It should be noted that, at the pH of 3.0 used here for SCX, the side chains of Asp and Glu are largely not charged unless the residue is next to a basic residue where the local microenvironment's pH would be higher. An extra basic residue at the N-terminus increased retention significantly. The increase in retention was slightly reduced when the extra basic residue was positioned at locations increasingly remote from the N-terminus but remained significant no matter where the basic residue was located. These data imply that the extra basic residue can function in cation exchange as a good alternative to the N-terminus as a binding site; orientation in this case is not as rigid. Location of the extra basic residue next to the Lys/Arg residue at the

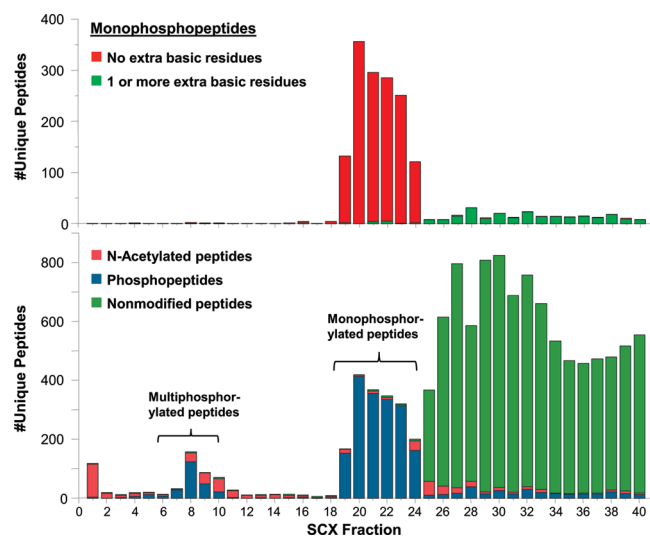


Figure 6. Elution in SCX of different classes of peptides in a Lys-N digest. Sample: Lys-N digest of HEK 293T cell lysate. Conditions: See Materials and Methods. (Bottom) Phosphopeptides that eluted in fractions 5–10 had no basic residues or had more than one phosphate group, while N-acetylated peptides eluting later than fraction 18 had extra basic residues. (Top) Distribution of monophosphopeptides with or without extra basic residues in addition to the N-terminal lysine.

C-terminus did not increase retention more than did its location anywhere else in the sequence, again implying that the C-terminus is not a favored binding site in SCX. The ANN results suggest broad trends regarding peptide orientation and the interaction of charged residues as a function of their position in the peptide.

Effect of a Phosphate Group's Position on Retention in SCX: Lys-N Digests. We recently reported that fractionation of peptides in a Lys-N digest using SCX can result in the sequential elution of specific classes of peptides, even when the classes carried the same net charge. Lys-N peptides acetylated at the N-terminus (and carrying the N-terminal basic residue) eluted just prior to peptides with a free N-terminus plus one phosphate group.^{18,19} Both classes of peptides have the same nominal net charge at the pH (2.7) used for the chromatography. The observation can easily be rationalized using the orientation model for peptide retention. If the N-terminus is preferred over the C-terminus as a binding site for a tryptic peptide, then that would emphatically be the case for a Lys-N peptide, where the N-terminus and the terminal Lys residue are located at the same end of the peptide. Acetylation of the N-terminus might then be expected to result in reduced interaction in SCX more reliably than would attachment of a phosphate group since a phosphate group located far from the N-terminus might have little effect on the chromatography if the peptide binding were oriented as shown in Figure 2 (right). The effect of phosphate location was assessed with a Lys-N digest of a lysate of HEK 293 cells. A histogram was constructed showing the distribution of singly phosphorylated peptides versus SCX fraction number (Figure 6, top). Most such peptides elute in a narrow range of six fractions if there are no basic residues besides the N-terminal Lys. Multiphosphorylated Lys-N peptides experience more electrostatic repulsion and elute

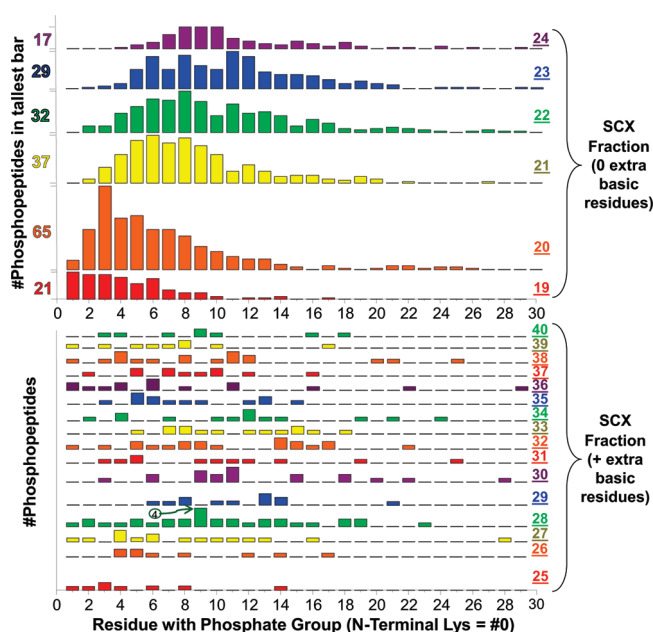


Figure 7. Lys-N phosphopeptides: Distribution in SCX fractions as a function of the distance of the phosphate group from the N-terminus. Sample and conditions: See Figure 6. (Top) Peptides with one phosphate group and a Lys residue at the N-terminus but no additional basic residues. The units in the Y-axis are the number of phosphopeptides in the tallest bar in each fraction. (Bottom) Peptides with one phosphate group and a Lys residue at the N-terminus plus at least one additional basic residue. The tallest bar (labeled) corresponds to four peptides. Both graphs share the same X-axis.

early along with the peptides acetylated at the N-terminus, as do peptides with one phosphate group and no basic residues (Figure 6, bottom). Monophosphopeptides with N-terminal acetylation that elute after fraction 24 have additional basic residues. The sequences and SCX fractions of elution of the Lys-N peptides are listed in the Supporting Information of ref 16. The histogram in Figure 7 (top) shows the distribution of singly phosphorylated peptides in SCX fractions 19–24 as a function of the distance of the phosphate group from the N-terminus. It is apparent that the farther the phosphate group is from the N-terminus, the better retained the peptide is in SCX, consistent with the orientation model in Figure 2 (right). This trend could potentially assist in confirmation of phosphosite assignments in peptides containing multiple Ser, Thr, or Tyr residues. This also accounts for the resolution of phosphorylation positional isomers in IEX and ERLIC, as was evident in Figure 1. Peptides with one phosphate group plus one or more additional basic residue (Figure 7, bottom) elute in a broad range of fractions to the end of the SCX gradient, with no discernible relationship between phosphate site and fraction of elution. This is consistent with the data in Figure 5 that indicated that an extra basic residue anywhere in the sequence may represent an alternative to the N-terminus as a stationary phase binding site in SCX, eliminating the rigid orientation of peptide binding observed for peptides containing just a single basic residue. The additional electrostatic attraction conferred by the residue varies as widely as does the location of the extra basic residue(s) in the sequences.

Effect of a Phosphate Group's Position on Retention in SCX: Tryptic Digests. Using the same SCX setup as described above, it was observed that tryptic peptides behave in a fashion

(19) Taouatas, N.; Altelaar, F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R. *Mol. Cell. Proteomics* **2009**, *8*, 190–200.

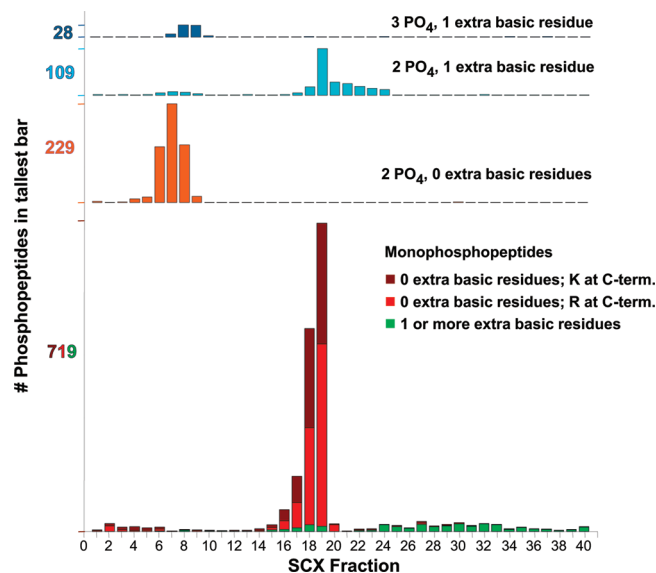


Figure 8. Tryptic phosphopeptides: effect of extra phosphate and basic residues on elution in SCX. Sample and conditions (see Figure 6): (Bottom) Peptides with one phosphate group and a Lys/Arg residue at the C-terminus, including peptides with additional basic residues and with segments corresponding to the number of peptides terminating in Arg or Lys residues. (Top) Peptides with additional phosphate groups and basic residues as indicated. The units in the Y-axis are the number of phosphopeptides in the tallest bar in each histogram.

similar to that observed for Lys-N peptides. Tryptic peptides with an acetylated N-terminus and one basic residue elute earlier than singly phosphorylated peptides with a free N-terminus and one basic residue even though both have a net charge of +1.¹⁸ Singly phosphorylated tryptic peptides with free N-termini elute in a narrower window in SCX than is true of singly phosphorylated Lys-N peptides (Figure 8, bottom). We observed no difference in elution times between singly phosphorylated tryptic peptides ending in Arg and those ending in Lys. As with Lys-N peptides, addition of additional basic residues leads to elution of phosphopeptides in much later fractions. Once again, the location of the phosphate group was plotted against the distribution of the peptides in the SCX fractions. Figure 9 (top) shows again a trend to later elution with increasing distance of the phosphate group from the N-terminus, although the trend is not as pronounced as with Lys-N peptides. Concomitantly, however, location of the phosphate group closer to the C-terminus does result in a more clear-cut trend to later elution (Figure 9, bottom). The phosphate group can hardly be contributing to retention at the C-terminal domain in SCX; rather, closer location to the C-terminus means that the phosphate group is farther from the N-terminus and less able to impair retention.

If a tryptic peptide has multiple phosphate groups and varying numbers of additional basic residues, then orientation effects appear to be abolished and elution is governed by the average charge of the peptide (Figure 8, top).

In line with our proposed model, the behavior of phosphorylated Lys-C peptides, also reported previously,¹⁸ was similar to that of tryptic peptides (data not shown). This is to be expected. A Lys-C fragment with the C-terminal Lys residue as the only basic residue is identical to a tryptic fragment and elutes in the same SCX fractions, while Lys-C fragments with additional Arg residues

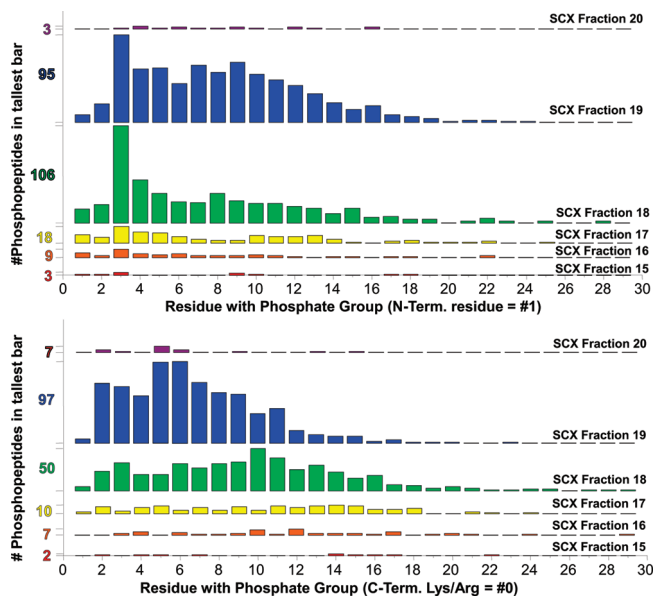


Figure 9. Tryptic phosphopeptides: distribution in SCX fractions as a function of the distance of the phosphate group from the N-terminus (top) or C-terminus (bottom). Sample and conditions: See Figure 6. The graphs include peptides with one phosphate group and a Lys/Arg residue at the C-terminus but no additional basic residues.

would be treated the same way here as tryptic fragments with additional basic residues. The main difference between the two digests was that the Lys-C digest had fewer peptides in the narrow-eluting cluster of peptides containing just one basic residue and more peptides in the range of peptides with additional basic residues.

DISCUSSION AND CONCLUSIONS

All of these data are consistent with a specific model of peptide binding in ion-exchange chromatography. If the peptide has one site or contact region that is significantly more favorable for binding than any other location in the peptide, then the peptide will tend to bind in an oriented fashion with that contact region down. The effect of other residues on retention will then be a function of their distance from the contact region. This model provides a rationale for selectivity in IEX of peptides and addresses a number of issues that have appeared in the literature:

- (1) Peptides with the same net charge can be separated by IEX: While factors such as charge-to-mass ratio surely play a role, the orientation effect explains why sequence is important as well. Any model for SCX elution rules should take this into account as well as net charge.
- (2) Tryptic or Lys-N peptides with a single phosphate group can be found throughout the gradient in SCX: Phosphopeptides with no extra basic residues elute in a tight cluster of fractions. The ones that elute later all have additional basic residues. The ones that elute earlier lack the terminal Lys/Arg residue.
- (3) Peptides with a blocked N-terminus elute in SCX before phosphopeptides with the same net charge: The N-terminus is the preferred binding site, whereas the phosphate group can be spaced well back in the sequence in a position where it has less influence on retention.

Orientation effects have been invoked in the literature²⁰ to account for the behavior of proteins in chromatography, particularly the concept of a preferred “contact region” for binding.^{21,22} Mant et al. compared the separation of amphipathic helical peptides in RPC and SCX-HILIC and found that RPC was sensitive to substitutions in the hydrophobic face of the peptide while SCX-HILIC was sensitive to substitutions in the hydrophilic face.²³ This was explained in terms of preferential binding of the face containing the substitution. The current study is not the first time that pronounced orientation effects have been demonstrated in chromatography of small solutes. It has been reported that residues near the termini of peptides contribute less to retention than do the same residues in the interior.⁵ Apparently, the termini are not preferred contact regions in RPC. The distribution of residues also affects retention; a set of hydrophobic residues promoted retention in RPC more when distributed throughout a peptide’s sequence than when they were in a block, either in the interior or at a terminus.²⁴ In a study of HILIC of carbohydrates,²⁵ disaccharides containing an amidated and a nonamidated sugar residue were always oriented with the amidated residue down. In that case, orientation was governed by polarity effects, amide groups being more polar than hydroxyl groups. It should be noted that the current results are contrary to those from an ANN study of anion exchange of tryptic peptides.²⁶ There, the position of a charged residue was not found to exercise any particular effect on the retention of a peptide. Of course, if the charged residue in question was an additional acidic one, then that would indeed abolish orientation effects and sensitivity to location in anion exchange (by analogy with Figures 6, 7, and 8).

In “bottom-up” proteomics analysis of complex samples, most peptides identified are the only ones identified from the protein in question. Since sequence identifications are frequently provisional, identification via a single fragment is generally not accepted without additional information that substantiates the identification of the peptide. “SCX elution rules” have been formulated to correlate the net charge of the peptide with its elution time in

SCX.¹⁰ Deviation from the expected elution window triggers a rejection of the putative sequence identification, thereby decreasing the number of false positives. There is frequently some uncertainty in determining which residue a phosphate group is located on, especially with multiple phosphate groups and multiple alternative sites of phosphorylation. If the same peptide is identified in different SCX fractions and with a phosphate group located at different positions, there is a tendency to ascribe this as a redundant identification because of the uncertainty inherent in current mass spectrometry methods for specifying the position of phosphate groups. The data in this paper indicate that some phosphorylation positional variants can readily be separated to the point that they elute in nonadjacent fractions and provide a rationale for declaring which variant should elute in which order. Additional rules can now be formulated to correlate the sequence with the elution time, especially in the case of phosphopeptides. A tentative list of rules might include the following:

- (1) A Lys-N, tryptic, or Lys-C fragment with a single phosphate group and no extra basic residues must elute in a narrow range in SCX along with all other peptides with those features. A peptide eluting later than this range should have a putative sequence containing at least one extra basic residue. A peptide eluting before this range should have a putative sequence missing the basic residue at a terminus or with the N-terminus blocked.
- (2) A Lys-N, tryptic, or Lys-C fragment with those features and where the phosphate group is within three residues of the N-terminus will elute in the early portion of that narrow range in SCX.
- (3) When positional isomers of phosphorylation are identified, then the variant with phosphorylation farthest from the N-terminus should elute last in SCX. A corollary is that such variants will also elute last in anion exchange and ERLIC because the phosphate group is then closer to the C-terminus.

Positional variants of phosphopeptides whose elution is consistent with these rules should not automatically be categorized as redundant identifications; they may be real variants whose existence is due to some variation in kinase positional specificity. We expect that further progress in this field will involve development of peptide retention time prediction algorithms as per the examples of Resing et al.⁹ and Petritis et al.,²⁷ who used the SCX retention time information to increase confidence in (unmodified) peptide annotations and to eliminate false-positive identifications.

Received for review March 11, 2010. Accepted May 6, 2010.

AC100651K

(20) Chung, W. K.; Hou, Y.; Freed, A.; Holstein, M.; Makhatadze, G. I.; Cramer, S. M. *Biotechnol. Bioeng.* **2009**, *201*, 869–881.

(21) Regnier, F. E. *Science* **1987**, *238*, 319.

(22) Xu, W.; Zhou, H.; Regnier, F. E. *Anal. Chem.* **2003**, *75*, 1931–1940.

(23) Mant, C. T.; Litowski, J. R.; Hodges, R. S. *J. Chromatogr. A* **1998**, *816*, 65–78.

(24) Alpert, A. J. *J. Chromatogr.* **1988**, *444*, 269–274.

(25) Alpert, A. J.; Shukla, M.; Shukla, A. K.; Zieske, L. R.; Yuen, S. W.; Ferguson, M. A. J.; Mehlert, A.; Pauly, M.; Orlando, R. *J. Chromatogr. A* **1994**, *676*, 191–202.

(26) Oh, C.; Žak, S. H.; Mirzaei, H.; Buck, C.; Regnier, F. E.; Zhang, X. *Bioinformatics* **2007**, *23*, 114–118.

(27) Petritis, K.; Kangas, L. J.; Jaitly, N.; Monroe, M.; Lopez-Ferrer, D.; Maxwell, R. A.; Mayampurath, A. M.; Petritis, B. O.; Mottaz, H. M.; Lipton, M. S.; Camp, D. G.; Smith, R. D. Strong cation exchange LC peptide retention time prediction and its application in proteomics, Poster# WP 591, *56th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO, June, **2008**.