Anal. Chem. **2005,** *77,* 2745−2755

Protein p*I* Shifts due to Posttranslational Modifications in the Separation and Characterization of Proteins

Kan Zhu,† Jia Zhao, and David M. Lubman*

Department of Chemistry, The University of Michigan, Ann Arbor, Michigan 48109-1055

Fred R. Miller

Karmanos Cancer Institute, Wayne State University, Detroit, Michigan 48202

Timothy J. Barder

Eprogen, 8205 South Cass Avenue, Suite 111, Darien, Illinois 60561

Proteins from breast cancer cell lines are characterized using a 2-D liquid separation technique in which protein p*I* is used as the first-dimension separation parameter. To effect this protein p*I* separation, chromatofocusing(CF) is employed whereby a pH gradient is generated on-column using a weak anion exchange medium with the intact proteins fractionated and collected every 0.2 pH unit. It is demonstrated that the p*I* for expressed intact proteins as generated by CF is an important parameter for identification and characterization of the actual protein modifications occurring in the cancer cell. For most proteins, the experimentally determined p*I* is very close to that predicted by the databases. In other cases, however, where the p*I* is observed to be shifted from the expected value, it is shown that this shift is often correlated to protein modifications. The modifications that cause such shifts include truncations and deletions often observed in cancer cells or phosphorylations that can shift the p*I* by several pH units. It is also shown that the effects of phosphorylation on the observed shift can vary depending upon the protein and the amount of phosphorlyation. Moreover, large changes in the p*I* are often observed for proteins with a p*I* above 7.0 upon phosphorlyation, whereas little change is observed for proteins with a p*I* of ~5.0. The expressed protein's p*I* value thus becomes an important parameter together with the intact MW value, peptide map, and MS/MS results for identification of the presence and type of posttranslational modifications occurring in the cancer cell.

An important issue in modern proteomics is the capability to map the protein expression of cells. This mapping has traditionally been performed using 2-D SDS−PAGE in which proteins are separated first in the pH dimension followed by MW separation using polyacrylamide gel electrophoresis. The separation of proteins based upon their isoelectric point (p*I*) has been an essential aspect of this separation, where p*I* is an inherent property of all proteins. The p*I* point occurs at a pH where the protein becomes neutral, and this property depends on the amino acid composition of the protein. Various algorithms have been proposed to predict the p*I* value from the amino acid composition.[1−3] Excellent agreement between the theoretical p*I*s and the experimental p*I* values obtained from isoelectric focusing has been achieved.[4−6] The use of pH is thus an important parameter for separation of proteins that is orthogonal to a second dimension based on the molecular weight of the protein. The use of p*I* becomes essential for separation of various modified isoforms of proteins that might be otherwise difficult to distinguish. More importantly, p*I* information has biological significance in proteomics as it is a physical property listed in the databases and is used in protein identification. Furthermore, the p*I* of eukaryotic proteins may reveal the subcellular localization.[7]

In recent work, there has been an effort to develop gel-free p*I*-based separations. These include liquid-phase isoelectricfocusing (IEF) using the Rotofor apparatus, which is a liquid-based analogue of carrier ampholyte gels,[8−12] and the IsoPrime device

* To whom correspondence should be addressed. E-mail: dmlubman@umich.edu.

† Current address: Novartis Institute for BioMedical Research, Inc., 250 Massachusetts Ave., Cambridge, MA 02139.

(1) Sillero, A.; Ribeiro, J. M. *Anal. Biochem.* **1989**, *179*, 319−325.
(2) Patrickios, C. S.; Yamasaki, E. N. *Anal. Biochem.* **1995**, *231*, 82−91.
(3) Halligan, B. D.; Ruotti, V.; Jin, W.; Laffoon, S.; Twigger, S. N.; Dratz, E. A. *Nucleic Acids Res.* **2004**, *32*, 638−644.
(4) Bjellqvist, B.; Hughes, G. J.; Pasquali, C.; Paquet, N.; Ravier, F.; Sanchez, J. C.; Frutiger, S.; Hochstrasser, D. *Electrophoresis* **1993**, *14*, 1023−1031.
(5) Bjellqvist, B.; Basse, B.; Olsen, E.; Celis, J. E. *Electrophoresis* **1994**, *15*, 529−539.
(6) Righetti, P. G.; Bossi, A. *Anal. Biochem.* **1997**, *247*, 1−10.
(7) Weiller, G. F.; Caraux, G.; Sylvester, N. *Proteomics* **2004**, *4*, 943−949.
(8) Kachman, M. T.; Wang, H. X.; Schwartz, D. R.; Cho, K. R.; Lubman, D. M. *Anal. Chem.* **2002**, *74* (8), 1779−1791.
(9) Nilsson, C. L.; Larsson, T.; Gustafsson, E. *Anal. Chem.* **2000**, *72* (9), 2148.
(10) Zhu, K.; Kim, J. K.; Yoo, C.; Miller, F. R.; Lubman, D. M. *Anal. Chem.* **2003**, *75*, 6209−6217.
(11) Wall, D. B.; Kachman, M. T.; Gong, S. Y.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. *Anal. Chem.* **2000**, *72* (6), 1099−1111.

10.1021/ac048494w CCC: $30.25 © 2005 American Chemical Society
Published on Web 03/24/2005

and other related units, which employ isoelectric membranes to separate proteins in the liquid phase and are the analogues of IPG-based gels.[13−17] Alternatively, continuous-flow electrophoresis has been used to separate proteins based upon p*I* in the liquid phase.[18,19] Capillary isoelectric focusing has also been used to separate proteins based upon p*I* in a one-dimensional separation interfaced to FT-ICR.[20] Although other methods, such as gel filtration,[21] ion exchange,[22] and size exclusion chromatography,[23] have been used as first dimensions in separations of proteins, these methods sacrifice the p*I* information as an important parameter in separating intact proteins and as a fundamental property in their identification.

More recently, chromatofocusing has been used as a column-based method for separating complex mixtures of proteins according to p*I*.[24−26] The CF technique depends on charge exchange on an ion exchange medium, where a pH gradient is generated by titration of a buffer against a start buffer, which sets the initial pH of the system.[27,28] In the case of weak anion exchangers, proteins are loaded on the column at a high pH, and as the titration proceeds, proteins with p*I*s greater than the pH will elute down the column, with high p*I* proteins eluting from the column first followed sequentially by proteins of successively lower p*I*s. Chromatofocusing has genuine advantages as a method for separating proteins based upon pH in that it can rapidly fractionate large numbers of proteins[25] and achieve separation of proteins in narrow (0.1) pH fractions. The pH can be measured on-line and the fractions collected directly as the eluent of the liquid-based separation. However, since CF is basically a charge exchange technique and not truly IEF, it may be questionable whether the p*I* measured is a valuable parameter for identification and characterization of a protein or whether it is simply a convenient separation tool.

In this work, we demonstrate that CF does indeed generate p*I* values for proteins similar to that predicted from the databases and that it is an important parameter for tagging the identity of the expressed intact protein. Moreover, CF can fractionate proteins into narrow pH ranges, which can be used to pin down the identity of proteins and to identify the presence of PTMs. In many cases where the measured p*I* does not match the database, it is shown that the protein is modified where the p*I* shift is consistent with these modifications. Further, the p*I* shift often is indicative of the number and type of modifications present.

(12) Hamler, R.; Zhu, K.; Buchanan, N. S.; Kreunin, P.; Kachman, M. T.; Miller, F. R.; Lubman, D. M. *Proteomics* **2004**, *4*, 562−577.
(13) Michel, P. E.; Reymond, F.; Arnaud, I. L.; Josserand, J.; Girault, H. H.; Rossier, J. S. *Electrophoresis* **2003**, *24*, 3−11.
(14) Zuo, X.; Speicher, D. W. *Proteomics* **2002**, *2*, 58−68.
(15) Zuo, X.; Speicher, D. W. *Anal. Biochem.* **2000**, *284* (2), 266.
(16) Zhu, Y.; Lubman, D. M. *Electrophoresis* **2004**, *25*, 949−958.
(17) Shang, T. Q.; Ginter, J. M.; Johnston, M. V.; et al. *Electrophoresis* **2003**, *24* (14), 2359−2368.
(18) Hoffman, P.; Ji, H.; Moritz, R. L.; Connelly, L. M.; Frecklington, D. F.; Layton, M. J.; Eddes, J. S.; Simpson, R. J. *Proteomics* **2001**, *1*, 807.
(19) Moritz, R. L.; Ji, H.; Schutz, F.; Connolly, L. M.; Kapp, E. A.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2004**, *76*, 4811−4824.
(20) Mohan, D.; Shen, Y.; Smith, R. D.; Lee, C. S. *Electrophoresis* **2002**, *23*, 3160.
(21) Meng, F.; Cargile, B. J.; Patrie, S. M; Johnson, J. R; McLoughlin, S. M.; Kelleher, N. L. *Anal. Chem.* **2002**, *74*, 2923.
(22) Liu, H.; Berger, S. J.; Chakraborty, A. B.; Plumb, R. S.; Cohen, S. A. *J. Chromatogr., B* **2002**, *782*, 267−289.
(23) Opiteck, G. J.; Ramirez S. M.; Jorgenson J. W.; Moseley, M. A., III. *Anal. Biochem.* **1998**, *258*, 349−361.
(24) Yan, F.; Subramanian, B.; Nakeff, A.; Barder, T. J.; Parus, S. J.; Lubman, D. M. *Anal. Chem.* **2003**, *75* (10), 2299−2308.
(25) Chong, B. E.; Yan, F.; Lubman, D. M.; Miller, F. R. *Rapid Commun. Mass Spectrom.* **2001**, *15* (4), 291.
(26) Zhu, K.; Miller, F. R.; Barder, T. J.; Lubman, D. M. *J. Mass Spectrom.* **2004**, *39*, 770−780.
(27) Sluyterman, L. A.; Elgersma, O. *J. Chromatogr.* **1978**, *150*, 17−30.
(28) Sluyterman, L. A.; Wijdenes, J. *J. Chromatogr.* **1978**, *150*, 31−44.

EXPERIMENTAL SECTION

MCF10CA1d.cl1 Cell Preparation and Lysis. The samples used in this work were prepared from a cloned malignant variant derived from the MCF10 series of human breast epithelial cell lines. Fully malignant human breast cancer cells, MCF10CAld clone 1 (CA1d) cells, were grown in monolayer on plastic in DMEM/F12 medium supplemented with 5% horse serum, 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES). Harvest was performed in the log phase (∼75%−80% confluence). The cells were gently washed with sterile PBS buffer before they were scraped with a rubber policeman and stored at −80 °C, after the growth medium was aspirated. A 1.0-mL sample of lysis buffer containing 6 M urea, 2 M thiourea, 0.5% (w/v) *n*-octyl β-D-glucopyranoside, 50 mM dithiothreitol, 2 mM tris(2−carboxyethyl)phosphine, 1 μL of protease inhibitor (Sigma, St. Louis, MO), and 10% (v/v) glycerol was added to cells prior to a 2-min vortex. The mixture was then left at room temperature for 1 h. To eliminate cell debris, sample and buffer mixture were centrifuged at 15 000 rpm for 20 min. The supernatant was collected, and the protein content was determined using a Bradford based assay (Bio-Rad, Hercules, CA).

Chromatofocusing. Lysis buffer in the whole cell lysate was replaced by the start buffer, which contained 25 mM Bis-Tris propane (Sigma) and 6 M urea. The pH was adjusted to 7.4 with saturated iminodiacetic acid (Sigma), using a PD-10 column (Amersham Pharmacia, Piscataway, NJ) according to the manual prior to loading the sample. Around 7 mg of protein extracted from CAld cells was loaded onto a CF column (Eprogen, Inc.) followed by elution at 0.2 mL/min using an elution buffer, a mixture of 10%(v/v) PolyBuffer 74 (Amersham Pharmacia) and 6 M urea with a pH value of 4.0. A saturated iminodiacetic acid was used for pH adjustment if required. Online pH measurement was performed as the eluent eluted from the column and before fraction collection using a pH electrode (Lazar Research, Los Angeles, CA) where the separation was monitored at 280 nm using a Beckman 166 model UV detector (Beckman-Coulter, Fullerton, CA). Eluent from the CF separation was collected from pH 7.4 to 4.0 every 0.2 pH unit intervals where 17 fractions were collected. All fractions were stored in a dry ice box after collection and transferred to a −80 °C freezer for storage.

Nonporous Silica (NPS) Reversed-Phase Liquid Chromatography. The NPS RP-HPLC separation was performed at a flow rate of 0.5 mL/min on an ODS IIIE (33 × 4.6 mm) column packed with 1.5-μm nonporous silica beads (Eprogen, Inc.). Each CF fraction contained more than 50 μg of protein, which was loaded onto the NPS RP column. During separation, the column was maintained at 65 °C with a column heater (Jones Chromatography model 7971) in order to improve the resolution and speed of the separation. Elution was performed using a water (A)/acetonitrile (B) (0.1% (v/v) trifluoroacetic acid, TFA) gradient. The gradient profile used was as follows: (1) 5%−15% B in 1 min, (2) 15%−25% B in 2 min, (3) 25%−31% B in 3 min, (4) 31%−41% B in 10 min, (5) 41%−47% B in 3 min, (6) 47%−67% in 4 min, (7) 67%−

100% in 1 min, (8) 100% B in 2 min, and (9) 100%−5% in 1 min. The acetonitrile was 99.93% HPLC grade (Sigma), and the TFA was from 1-mL sealed ampules (Sigma). The HPLC system was a Beckman model 127 HPLC where the separation is monitored at 214 nm using a model 166 detector. The proteins separated by NPS-HPLC were collected into 1.5-mL tubes using a Beckman SC-100 fraction collector controlled by in-house software. Purified samples were stored in dry ice before concentration using the speedvac.

NPS-RP-HPLC/ESI-TOF MS. Eluent from NPS-RP HPLC was analyzed on-line using ESI-TOF MS (LCT, Micromass, Manchester, U.K.) for MW of intact proteins. The separation was performed under the same conditions as in the previous section except for the addition of 0.3% formic acid (Sigma) in both mobile phases to improve ESI efficiency. The eluent was introduced into the ESI-TOF at a flow rate of 200 μL/min. The capillary voltage for electrospray was set at 3200 V, sample cone at 40 V, extraction cone at 3 V, and reflection lens at 750 V. Desolvation was accelerated by maintaining the desolvation temperature at 300 °C and source temperature at 120 °C. The nitrogen gas flow was controlled at 650 L/h. During the separation, one mass spectrum was acquired per second. The intact molecular weight value was obtained by deconvoluting the combined ESI spectra of the protein with MaxEnt 1 software (Micromass).

Peptide Preparation for MS Analysis. Prior to tryptic digestion, purified proteins were dried down to 20 μL using a CentriVap concentrator (Labconco Corp., Kansas City, MO) followed by the addition of 20 μL of 100 mM ammonium bicarbonate (Sigma) to adjust the solution pH to ~7.8. A 0.5-μg aliquot of L-1-tosylamido-2-phenylethyl chloromethyl ketone-modified sequencing grade trypsin (Promega, Madison, WI) was added prior to vortexing. The mixture was stored in a warm room with a controlled temperature of 37 °C for 24 h where the samples were shielded from light. After digestion was halted by adding 1 μL of 10% (v/v) TFA, the tryptic digest mixture was concentrated to 5 μL using a ZipTip(Millipore, Billerica, MA).

MALDI-TOF MS. Sample spotting was performed by layering 1 μL of matrix on top of 1 μL of sample. The spot was allowed to dry in air. The MALDI matrix was prepared by diluting saturated α-CHCA (Sigma) solution made in 50% (v/v) ACN and 1% (v/v) TFA with the same solution at a 1:4 ratio (v/v). Internal standards were prepared as 1 mg/mL angiotensin I, ACTH 1−17, and ACTH 18−39 (Sigma) and were further diluted 100-fold with deionized water. Aliquots of 13, 21, and 25 μL were taken from each of the diluted standard solutions respectively and mixed with matrix to produce 1 mL of solution containing 50 fmol of each standard per spot.

Peptide masses were measured on a Micromass TofSpec2E (Micromass/Waters, Milford, MA) with delayed extraction in the reflectron mode using a nitrogen laser (337 nm). Peptide mass spectra were internally calibrated using the three peaks from internal standards resulting in a mass accuracy of 25 ppm or less. The calibrated spectra were processed using PeptideAuto (Micromass MassLynx application) to obtain experimental masses that were submitted to MS-Fit (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm) to search the SwissProt database for protein identity. The following parameters were used for database searching: maximum number of two missed cleavages, unmodified cysteine, *Homo sapiens* for the species, no restriction on the pH range, monoisotopic mass.

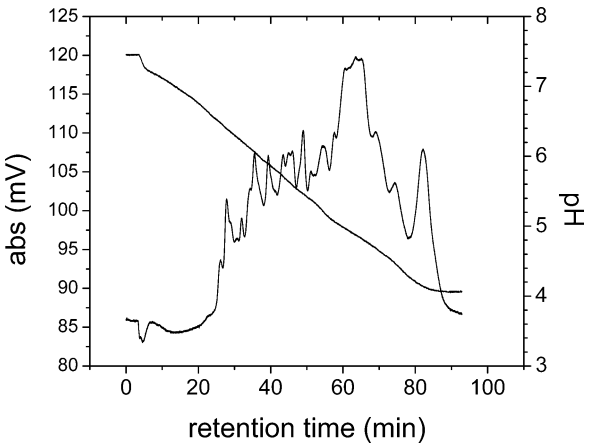

Figure 1. CF chromatogram (dark solid line) of 6.2 mg of Ca1DCL1 whole cell lysate. Elution is conducted at 0.2 mL/min and monitored at 280 nm using a UV detector. pH measurement was achieved online using a pH electrode connected with a flow cell after the UV detection. The linear pH gradient is also shown in the figure.

Capillary LC-ESI-QTOF. Each sample that was analyzed by MALDI TOF MS was subjected to further analysis using capillary LC-ESI-QTOF. Prior to LC separation, 50 μL of deionized water was added to each sample. Each of the diluted samples was further concentrated to ~5 μL using a SpeedVac so that the acetonitrile in solution was removed. All 5 μL of concentrated sample was loaded to an 180 μm × 150 mm capillary column packed with 3-μm C18 silica particles (PepMap, LC Packings, Sunnyvale, CA). Separation was conducted using an Ultimate capillary HPLC system (LC Packings, Sunnyvale, CA) which maintains a flow rate of 2 μL/min and generates a gradient of 5%−40% B in 60 min. In both mobile phases, 0.1% formic acid was added to enhance separation in LC and ionization efficiency in the electrospray. Furthermore, mobile phase A consisted of 0.1% (v/v) formic acid/deionized water and B was composed of 0.1% formic acid/ACN. Eluent from the capillary LC was interfaced through a nanospray source to a Qtof Ultima Global quadrupole time-of-flight mass spectrometer (Micromass). The positive MS survey scan was performed to monitor the separation until MS/MS was triggered when the intensity of the observed peptides exceed a set threshold value of 25 counts. Obtained MS/MS data were processed by PeptideAuto and searched against the NCBInr protein database using the MASCOT program (Matrix Science) where the mass tolerance was set at 0.2Da.

RESULTS AND DISCUSSION

Separation of Proteins in the pH Dimension. Chromatofocusing. Chromatofocusing (CF) is a protein separation technique that combines the high capacity of ion exchange chromatography with the ability to separate proteins on the basis of their p*I* values.[29] The separation of 6.2 mg of Ca1DCL1 breast cancer whole cell lysate using CF is shown in Figure 1. Around 12 broad peaks were observed when the pH was gradually decreased from

(29) Li, C. M.; Hutchens, T. W. *Pratical Protein Chromatography*; Kenney A., Fowell, S., Eds.; Methods in Molecular Biology 11; Humana Press: Totowa NJ. 1992; Chapter 15, p 237.

Table 1. Proteins Identified from CA1dCL1 Breast Cancer Cell Line over p*I* Range of 4.6−5.4

ID	acc no.	theor MW	theor p*I*	exp MW	exp p*I*
protein disulfide isomerase A4 precursor (ERp72)	P13667	72 933	5.0	72 731	5.0−5.2
GRP78	P11021	72 334	5.1	71 396	5.0−5.2
HSP70kDa protein	P08107	70 053	5.5	70 129	5.0−5.2
CK5	P13647	62 462	8.1	62 502	4.6−4.8
keratin, type I cytoskeletal 9 (cytokeratin 9) (K9) (CK 9)	P35527	61 988	5.1	61 883	4.8−5.0
ATP synthase α chain, mitochondrial precursor	P25705	59 751	9.2	59 752	4.8−5.0
HSP60	P10809	61 055	5.7	57 963	5.2−5.0
protein disulfide isomerase A3 precursor (disulfide isomerase ER-60) (ERp60) (58 kDa microsomal protein) (p58) (ERp57) (58 kDa glucose regulated protein)	P30101	56 783	6.0	56 696	5.0−5.2
vimentin	P08670	53 686	5.1	53 635	5.2−5.4
CK8	P05787	53 675	5.5	53 630	5.4−5.2
ATP synthase β chain, mitochondrial precursor	P06575	56 560	5.3	51 850	5.0−5.2
keratin, type II cytoskeletal 7 (cytokeratin 7) (K7) (CK 7) (sarcolectin)	P08729	51 418	5.5	51 387	5.4−5.2
keratin, type I cytoskeletal 16 (cytokeratin 16) (K16) (CK 16)	P08779	51 268	5.0	51 336	4.8−5.0
CK15	P19012	49 168	4.7	49 092	5.0−4.8
CK18	P05783	48 058	5.3	48 044	5.2−5.0
keratin, type I cytoskeletal 19 (cytokeratin 19) (K19) (CK 19)	P08727	44 106	5.0	44 066	5.0−5.2
CK6D	P48667	42 469	5.3	42 518	5.0−5.2
actin, cytoplasmic 2 (γ-actin)	P02571	41 793	5.3	41 792	5.2−5.4
annexin A1 (annexin I) (lipocortin I) (calpactin II) (chromobindin 9) (P35) (phospholipase A2 inhibitory protein)	P04083	38 715	6.6	38 679	5.2−5.4
G1/S-specific cyclin D1 (PRAD1 oncogene) (BCL-1 oncogene)	P24385	33 729	5.0	33 728	5.0−5.2
tropomyosin α 4 chain	P07226	28 522	4.7	28 441	4.6−4.8
splicing factor, arginine/serine-rich 9	Q13242	25 542	8.1	25 536	4.8−5.0
TCTP	P13693	19 596	4.8	19 598	4.8−5.0

pH 7.4 to 4.0 in 85 min. The use of continuous real-time postcolumn pH measurements demonstrates the linearity of the pH gradient, as shown in the gradient line in Figure 1. During the CF fractionation, proteins were collected every 0.2 pH unit change, which is controlled by computer so that adjacent fractions differ in pH by 0.2 pH unit. Each of the fractions with pH values within pH 5.4−4.6 was separated by NPS-RP-HPLC, and the eluent was analyzed on-line by ESI-TOF MS for the MW value of the intact proteins.

The separations obtained in CF are similar to those obtained in the electrophoretic method in that proteins are separated according to their p*I*s. While the pH gradient is linear in CF, neither capillary isoelectric focusing nor IPG gel electrophoresis achieves a linear pH gradient.[30] The result is that isoelectric point markers must be introduced to derive the relation between the migration distance and the pH value. In contrast, a true online pH measurement is achieved in CF. Further, no protein is subjected to a pH higher than approximately its own p*I* value so that the separation condition is mild. Since proteins remain in the solution phase during separation, sample recovery for postseparation analysis is readily achieved.

The net charge on a protein is determined by the pH of the environment and the p*I* of the protein. A protein is negatively charged when its p*I* is lower than the pH of the environment, while the protein will be positively charged when its p*I* is higher than the environmental pH. In these CF experiments, all proteins with p*I* lower than 7.4 will be negatively charged and be retained by the positively charged stationary phase. The introduction of the elution buffer, which has a pH value of 4, changes the charge on the protein by altering the pH of the environment, hence the interaction of the charged protein with the stationary phase. In contrast to electrophoretic focusing techniques, separation by chromatofocusing relies on interfacial (solid/solution phase) molecular electrostatic interaction, which is affected by ionic strength, competitive binding reagents (displacers), nonspecific adsorption properties of the stationary phase and the distribution of the surface charge of proteins.

Proteins Elute at Their p*I*s. Theoretical analysis conducted by Sluyterman et al. indicated that the apparent p*I* values obtained from CF approximate the p*I* values observed in the electrophoretic method.[27,28] In our results, the experimental p*I* of many proteins match the theoretical p*I* of the corresponding proteins, as shown in Table 1. The protein identifications in Table 1 were performed using MALDI-TOF MS peptide mapping and the MW value of the intact protein as obtained using ESI-TOF MS. In addition, ESI-MS/MS using the QTOF was used to verify the identification, especially where any ambiguity existed. While theoretical p*I*s were obtained based upon protein identification using peptide mapping and matching against the databases, all experimental p*I*s in the table were obtained from the real-time pH measurement of the collected CF fraction during CF elution. In the table, most proteins are present in one CF fraction so that the experimental p*I* equals the measured pH value of the CF fraction. However, some abundant proteins may be present in more than one CF fraction. The distribution of three abundant proteins in CF fractions with various pH values is shown in Figure 2. Both CK18 and HSP60 were detected in three CF fractions, and CK15 was observed in four fractions. The presence of these proteins in the corresponding CF fraction has been verified using both MALDI-TOF MS peptide mapping and the MW of the intact proteins obtained from deconvolution of the ESI-TOF MS spectrum acquired from the NPS-RP-LC-ESI-TOF MS analysis. The peak area of the deconvoluted ESI-TOF MS spectrum reflects the abundance of each protein in the corresponding CF fraction since the NPS-RP-HPLC-

(30) Shimura, K.; Zhi, W.; Matsumoto, H.; Kasai, K. *Anal. Chem.* **2000**, 4747−4757.

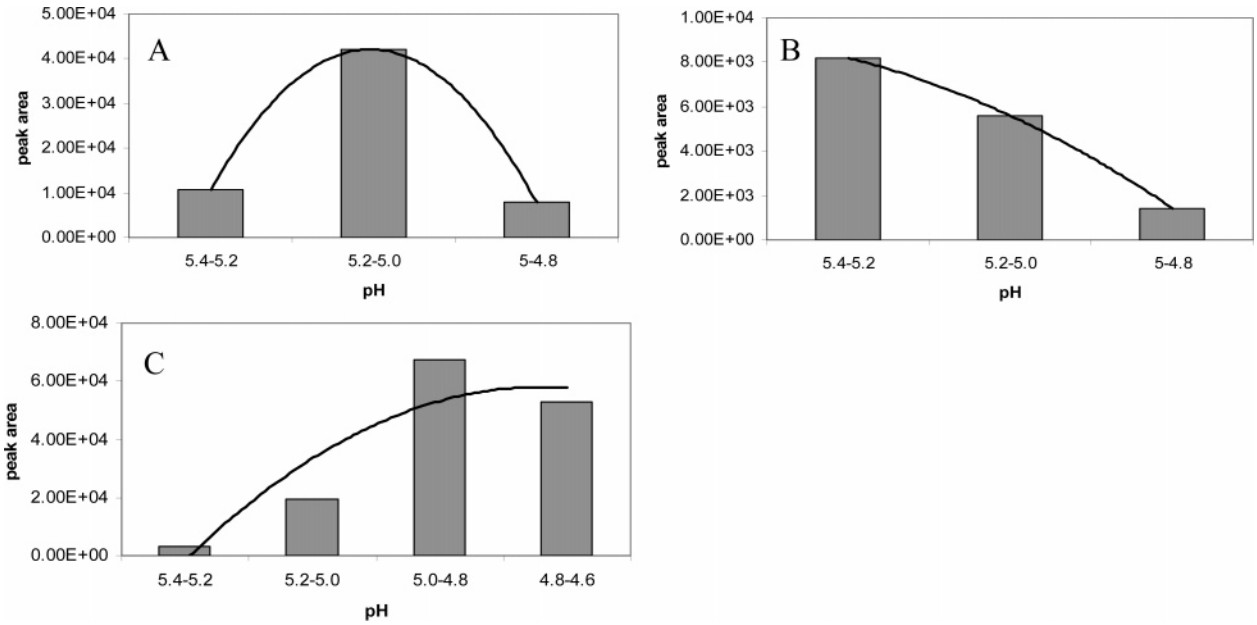

Figure 2. (A) CK18 (p*I*, 5.3), 87.1% of CK18 focused to pH 5.4−5.2 and pH 5.2−5.0. (B) HSP60 truncated (p*I*, 5.24), 90.8% HSP60 was focused to pH5.4−5.2 and pH 5.2−5.0. The precursor of HSP60 has a p*I* of 5.7. (C) CK15 (p*I*, 4.7), 84.2% of CK15 was focused to pH 5.0−4.8 and pH 4.8−4.6.

ESI-TOF MS analysis has been conducted under the same experimental conditions. Curve fitting using the quadratic function shows the maximum abundance of CK18 at pH 5.2−5.0, HSP60 (truncated) at pH 5.2−5.0, and CK15 at pH ∼4.8. It should be noted that the amount of CK18 focused at pH 5.4−5.2 and pH 5.2−5.0, HSP60 (truncated) at pH 5.4−5.2 and pH 5.2−5.0, and CK15 at pH 5.0−4.8 and pH 4.8−4.6 account for 87.1%, 90.8%, and 84.2%, respectively, of the respective total amount of protein in the whole cell lysate. Thus, we can conclude that high-abundant proteins were focused to two CF fractions with pH values closest to their corresponding p*I*s. The incomplete focusing of these abundant proteins is due to the saturation of binding sites on the stationary phase. A similar band-broadening effect is common in other LC separation techniques such as reversed-phase LC and ion exchange chromatography. To observe low-abundant proteins, the large amount of sample loaded typically overloads the CF column for high-abundant proteins. Nevertheless, our result indicates that focusing is still achievable in CF. Furthermore, the 24 selected proteins in Table 1 vary in size and molecular weight. Our results support the idea that the separation of proteins using CF is based on the p*I* of proteins. It should be noted that 6 M urea was used in both the start buffer and the elution buffer so proteins remain denatured during separation.

Deviation of the Experimental p*I* from the Corresponding to Theoretical Values. CF separates proteins according to their p*I* values; however, deviation of the experimental p*I* from its corresponding theoretical value does occur. Some deviation of the p*I* may result from the interaction of a partial protein sequence with the stationary phase. However, most deviations in p*I* appear to be due to posttranslational modifications. For example, the first 26 amino acids of HSP60 mitochondrial precursor have been reported to be truncated in our previous work.[10] The truncation results in a decrease of p*I* of ∼0.5 pH unit in CF. Several additional examples are shown in Table 2, in which all selected proteins present discrepancies between the experimental and theoretical p*I*. In ATP synthase coupling factor 6 mitochondrial precursor, the deletion of the first 32 amino acids results in the highest shift of 4 pH units, where the theoretical p*I* of pH 9.5 is reduced to 5.5. In comparison, the truncation of the first 47 amino acids of ATP synthase β chain (mitochondrial precursor) only slightly decreased its theoretical p*I* by 0.3 pH unit. In Table 2, the p*I* shift of the first five proteins due to truncation agrees quite well with the corresponding experimental p*I*. Among the 20 amino acids, three basic amino acids—lysine (K), arginine (R), and histidine (H)—can bear positive charge, and four acidic amino acids—tyrosine (Y), cysteine (C), aspartate (D), and glutamate (E)—can be negatively charged.[7] The shift of the theoretical p*I* is due to the change in the total number of basic and acidic residues in a protein after sequence deletion.[6]

The truncation of these proteins is further confirmed by the experimental mass values. These proteins vary in size from 9 to 60 kDa, which suggests that the separation according to the p*I* of proteins is not affected by the size of proteins. Further, the detection of these proteins in the truncated form reveals that they were extracted from mitochondria.[31] The experimental p*I* information does help to localize these proteins in cells.

Experimental p*I* can be applied to distinguish protein isoforms. ATP synthase D chain mitochrondrial has two isoforms, isoform 1 and isoform 2. These proteins are different splicing products where residues 73−96 are missed in isoform 2. The theoretical p*I* of isoform 2 is 6.86, while isoform 1 has a p*I* of 5.2. The observed p*I* of 5.2−5.0 suggests that isoform 1 is present in the sample. The experimental MW value of ATP synthase D chain mitochondrial, which is 18 405, confirmed that isoform 1 is present in this breast cancer sample. Further, the experimental MW of isoform1 is 45 higher than the theoretical value, which indicates acetylation. The N-terminal of the protein has been reported to be acetylated.[32] It should be noted that the experimental p*I* did

(31) Neupert, P. *Annu. Rev. Biochem.* **1997**, *66*, 863−917.

Table 2. p*I* Shift for Selected Proteins

protein	acc no.	theor MW	theor p*I*	modification	MW[a]	p*I*[a]	exp MW	exp p*I*
60 kDa heat shock protein, mitochondrial precursor (Hsp60) (60 kDa chaperonin) (CPN60) (heat shock protein 60) (HSP-60) (mitochondrial matrix protein P1) (P60 lymphocyte protein) (HuCHA60)	P10809	61 055	5.7	1−26 AAs truncated	57 966	5.24	57 963	5.4−5.2
ATP synthase β chain, mitochondrial precursor	P06576	56 560	5.3	1−47 AAs truncated	51 769	5.00	51 764	4.8−5.0
NADH-ubiquinone oxidoreductase 23 kDa subunit, mitochondrial precursor	O00217	23 705	6.0	1−34 AAs truncated	20 290	5.1	20 182	4.8−5.0
thioredoxin	Q99757	18 383	8.5	1−59 AAs truncated	11 867	4.9	11 868	5.0−4.8
ATP synthase coupling factor 6, mitochondrial precursor	P18859	12 588	9.5	1−32 AAs truncated	8 960	5.5	8 962	5.4−5.2
ATP synthase D chain, mitochondrial	O75947	18 360	5.2	isoform 2	15 642	6.86	18 405	5.2−5.0

[a] Values in this column were calculated based upon truncated proteins or isoforms.

not show a shift from the theoretical value, which suggests that acetylation does not induce changes in p*I*, or such changes are too small to be detected by CF.

Effect of Phosphorylation on the p*I* of Proteins. (1) Proteins with Slight p*I* Shift upon Phosphorylation. Protein phosphorylation is a key regulatory mechanism in cells.[33−35] This process affects the p*I* of intact proteins by substituting the neutral hydroxyl groups on serine, threonine, or tyrosine residues with negatively charged phosphate group(s).[3] As a result, phosphorylation usually induces an acidic shift on p*I*. It has been reported that one phosphorylation may decrease the p*I* by 1−2 pH units,[36] where multiple phosphorylations in one protein may result in a significant p*I* decrease. Although such results have been observed in 2-D gel experiments,[36] the same result should be observed in the use of CF despite the difference in the separation mechanism. However, our detailed experiments have shown that such shifts in p*I* may occur in some cases but not in others.

Vimentin is one of the proteins listed in Table 3. The p*I* of its intact form has a theoretical p*I* value of 5.1. Vimentin was isolated from the breast cancer cell lysate using the 2-D liquid fractionation. When the tryptic digest of this protein was analyzed using capillary LC-ESI-QTOF MS, 19 peptides were sequenced and assigned to vimentin. The corresponding base peak ion chromatogram is shown in Figure 3. Among the 19 peptides, SLYAS*p*SPGGVYATR was shown to be phosphorylated where the peptide elutes at 23.74 min. The observation of the y_9^{2+}-49 in the MS/MS spectrum in the inset of Figure 3 narrows the phosphorylation site down to S56, which is the sixth residue from the N-terminus of the peptide. The phosphorylation on serine barely affects the p*I* of the intact vimentin since the experimental p*I* matches well with the theoretical p*I* of its unmodified intact form. Since 6 M urea was used in both sample introduction and elution in CF fractionation, proteins should be completely denatured. As a result, all charged amino acids contribute to the p*I*, which is one of the assumptions in calculating the theoretical p*I* of proteins according to their primary sequence. In addition, the theoretical MW with one phosphorylation does not match the experimental value, indicating the presence of additional modifications, which as yet have not been identified.

Other proteins in Table 3 include CK17, hepatoma-derived growth factor, CK15, and CK8. All of these proteins have one or more detected phosphorylations. Moreover, the measured p*I* values agree well with the theoretical p*I* values of the corresponding phosphorylated form where no significant shifts were observed. Similar results have been observed on acetylcholine receptor,[37] a protein with p*I* of ∼5.4. After treatment of its phosphorylated form with alkaline phosphatase, a shift of 0.1 pH unit to a more alkaline region was observed.

(2) Theoretical p*I* for Phosphorylated Proteins. Recently, software has been developed to predict the effect of phosphorylation on the p*I* of a protein.[3,38] One such program is the protein modification screening tool (ProMoST) developed by Halligan et al.[3] Phosphorylation induces an acidic shift on p*I* by introducing negative charge(s). The shift of p*I* for proteins with multiple phosphorylations is cumulative for each phosphorylation. In this work, ProMoST is used to calculate the p*I* for each of the phosphorylated proteins in Table 3. The number of phosphorylations used in calculation is obtained by combining all the detected phosphorylations for a particular protein. As shown in Table 3, both cytokeratin 17 and hepatoma-derived growth factor have four phosphorylations, so their access number and four phospho Ser/Thr are provided to calculate their theoretical p*I*. The change in p*I* due to phosphorylation is obtained by subtracting the calculated p*I* of the phosphorylated proteins from the p*I* of the nonphosphorylated proteins. As shown in Figure 5, the change of p*I* is plotted against the number of phosphorylations. The figure reveals that, even with up to 10 phosphorylations (ST), the p*I* change is still less than 0.65 pH unit (CK8). In some cases,

(32) http://us.expasy.org/cgi-bin/niceprot.pl?O75947.
(33) Hunter, T. *Cell* **1995**, *80*, 225.
(34) Faux, M. C.; Scott, J. D. *Trends Biol. Sci.* **1996**, *19*, 480.
(35) Gschwind, A.; Fischer, O. M.; Ullrich, A. *Nat. Rev. Cancer* **2004**, *4* 361−370.
(36) Yamagata, A.; Kristensen, D. B.; Takeda, Y.; Miyamoto, Yuka; Okada, K.; Inamatsu, M.; Yoshizato, K. *Proteomics* **2002**, *2*, 1267−1276.
(37) SaïToh, T.; Changeux, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78* (7), 4430−4434.
(38) Kumar, Y.; Khachane, A.; Belwal, M.; Das, S.; Somsundaram, K.; Tatu, U. *Proteomics* **2004**, *4*, 1672−1683.

Table 3. Proteins with No p*I* Shift upon Phosphorylation

protein	acc no.	MW	peptide sequence	phospho (at protein)	obsd phos sites[a]	theor p*I* database[b]	exp p*I*	theor p*I*[c] calcd	pred phos sites[d]
vimentin	P08670	53 686	SLYAS*p*S*p*SPGGVYATR	S55	1	5.1	5.2−5.4	5.05	Ser 16 Thr 1 Tyr 1
cytokeratin 17 (*H. sapiens*)	Q04695	48 106	QFTSSS*p*SIK	S12	4	5.0	5.0−4.8	4.83	Ser 9
			L*p*SGGLGAGSCR	S31					Thr 1
			QF*p*T*p*SSSSIKGSSGLGGGSSR	T8, S9					Tyr 2
hepatoma-derived growth factor (high-mobility group protein 1-like)	P51858	27 727	GNAEG*p*S*p*SDEEGKLVIDEPAK	S132, S133	2	4.7	4.8−4.6	4.58	Ser 7
hepatoma-derived growth factor			KGNAEG*p*S*p*SDEEGKLVIDEPAK	S132, S133					Thr 1
keratin, type I cytoskeletal 15 (cytokeratin 15) (K15) (CK 15)	P19012	50 360	FVSSG*p*SGGGYGGGMR	S56	1	4.7	4.8−4.6	4.69	Ser 10
keratin 8; cytokeratin 8; keratin, type II cytoskeletal 8	P05787	54 666	I*p*SSSSFSR	S34	2	5.5	5.4−5.2	5.44	Thr 1 Tyr 2 Ser 25
			TTSGYAGGLSSAYGGL*p*TSPGL-SYSLGSSFGSGAGSSSFSR	T431					Tyr 4

[a] The total number of phosphorylation sites observed in the LC−MS/MS analysis for each protein. [b] Theoretical p*I* of each intact protein in its unmodified form. [c] The p*I* of proteins in the phosphorylated form. They are calculated using ProMoST (http://prometheus.brc.mcw.edu/promost/) when the total number of detected phosphorylations and the access numbers of the protein are provided. 1, 4, 2, 1, and 2 phosphorylations are used in calculating the p*I* for vimentin, CK17, HDGF, CK15, and CK8, respectively. [d] The predicted phosphorylation sites are calculated using the NetPhos software (http://www.cbs.dtu.dk/services/NetPhos/). The FASTA sequence for each of the listed proteins is used as input. However, only the phosphorylation sites with score greater than 0.9 are reported.

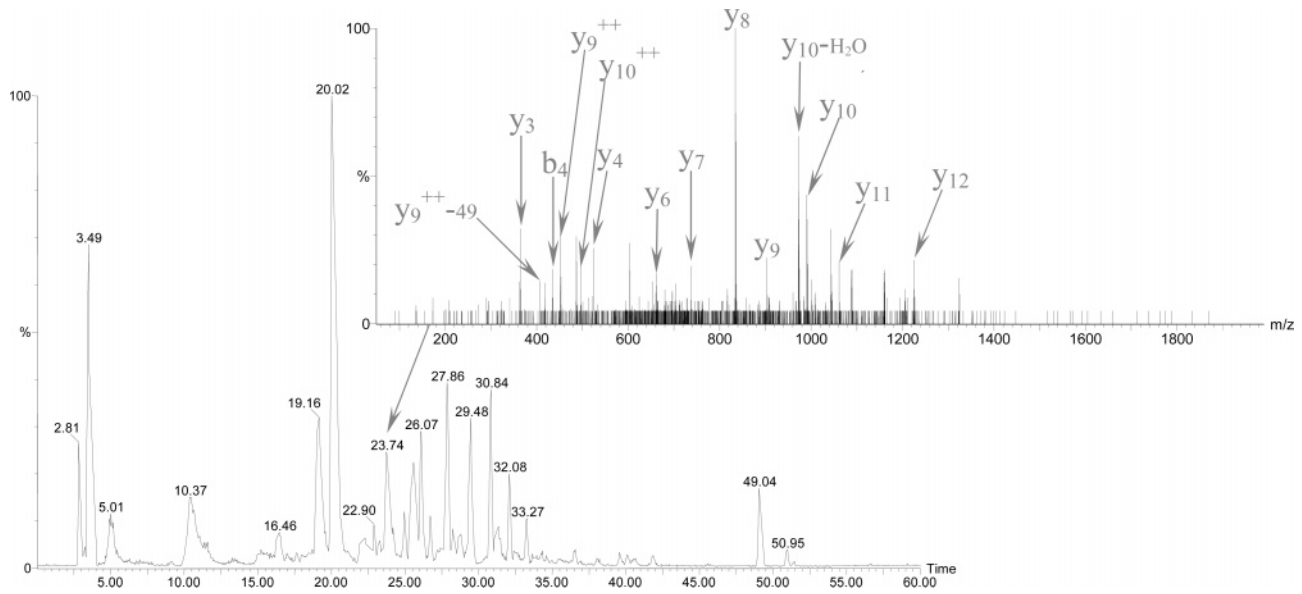

Figure 3. Base peak ion chromatogram of the capillary LC-QTOF survey scan experiment for the tryptic digest of vimentin. The inset is the MS/MS spectrum of SLYAS*p*SPGGVYATR(2+). The separation is conducted on an 180 μm × 150 mm C18 column packed with 3-μm silica particles with pore size of 300 Å. A linear gradient of 5%−40% B in 60 min is applied, and 2 μL/min flow rate is used. MS/MS acquisition is triggered when the ion count exceeds 25 counts.

the shift is even lower than 0.3 pH unit (CK15). However, as shown in Table 3, the calculated p*I* values agree well with the experimental p*I*s when the number of detected phosphorylations is used in calculation for each protein. In the case of CK15, phosphorylation barely induces a p*I* shift. In Figure 5, it is shown that the effect of phosphorylation on the p*I* of proteins is protein dependent and very much dependent on the number of phosphorylations. For example, CK8 has a nonlinear curve where the first several phosphorylations induce a greater shift than the last several phosphorylations. Other proteins have somewhat linear

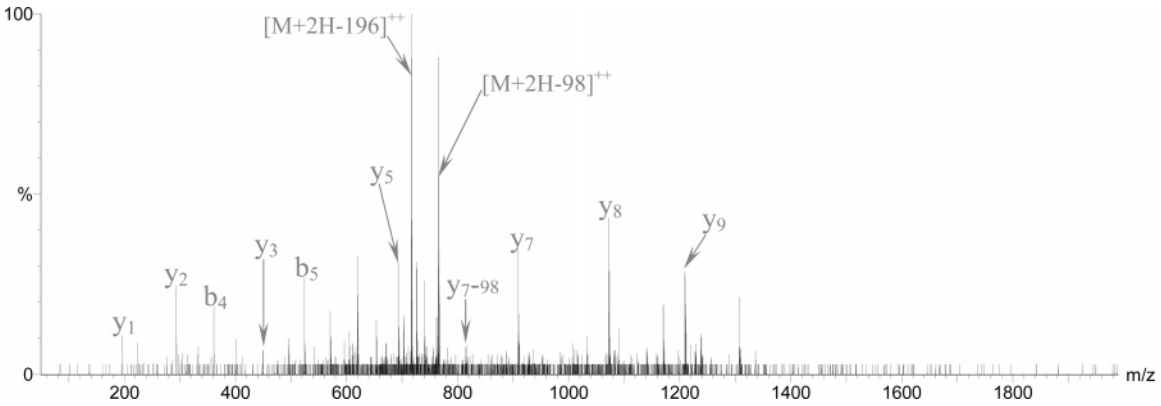

Figure 4. MS/MS spectrum of G*p*SPHYF*p*SPFRPY (2+) from splicing factor, Arg/Ser-rich 9.

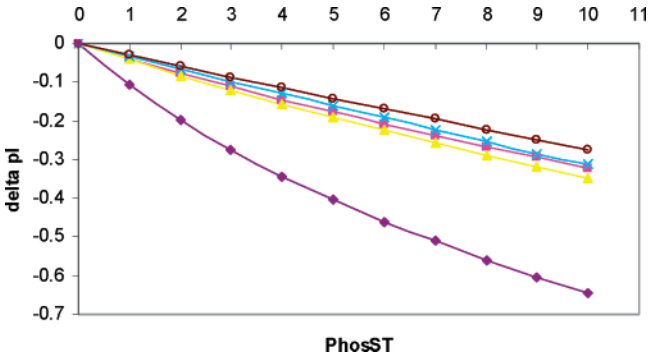

Figure 5. Variation of the theoretical p*I* values as a function of the number of phosphorylations on vimentin (■), CK17 (▲), HDGF (×), CK15 (○), and CK8 (◆). The theoretical p*I* values are calculated using the ProMoST.

curves, which indicates that each phosphorylation induces the same value of shift in p*I*. This analysis is based on the assumption that these phosphorylation sites are considered discrete and no interactions between phosphorylations were considered. It should be noted that all these proteins have p*I* values close to pH 5.0.

(3) Prediction of Phosphorylation Sites According to the Primary Sequence of Proteins. Protein phosphorylation sites can be predicted from protein sequence and structure.[39] Protein kinases are a family of enzymes that catalyze phosphorylation events. They share homologous catalytic domains and similar mechanisms of substrate recognition.[40−43] Acidic residues are usually found several residues away from the tyrosine phosphorylation sites. Also, the structure close to the phosphorylation site should be flexible, while for serine/threonine phosphorylation sites, a basic motif is usually observed. Blom et al.[39] developed an artificial neutral network method to predict phosphorylation sites according to the primary sequence. The sensitivity of the method ranges from 69% to 96%. In this study, we used the method to predict the likely phosphorylation sites for the phosphorylated proteins observed in our experiments.

To predict phosphorylation sites, the protein sequence in FASTA format was used as input for NetPhos 2.0 software (http://www.cbs.dtu.dk/services/NetPhos/). Only sites with scores greater than 0.9, which suggests very likely phosphorylation sites, were reported. As shown in Table 3, vimentin has 16 serine, 1 threonine, and 1 tyrosine phosphorylation sites. However, only one phosphorylation on serine has been detected in our LC−MS/MS experiments. Similar results were also observed in CK8 and CK15 where there are 29 and 13 predicted phosphorylation sites but only 2 and 1 serine phosphorylations were detected, respectively. Alternatively, a third of the phosphorylation sites in CK17 and half of the phosphorylation sites in HDGF were confirmed as phosphorylated. The corresponding detected phosphorylated peptides in LC−MS/MS analysis are shown in Table 3. The deviation of the number of phosphorylations as determined by LC−MS/MS from that predicted by the software may result from either error in calculation or incomplete coverage from LC−MS/MS. Some of the predicted sites may not be real due to the incompleteness of the theoretical model, while some real phosphorylation sites remain unidentified due to the limitation of the LC−MS/MS experiments. These limitations result from suppression effects in ionization, limited MS/MS speed, and low abundance of phosphorylated peptides, especially for peptides with phosphotyrosine.

(4) Proteins with Significant p*I* Shift upon Phosphorylation. In contrast to the minor p*I* decrease upon phosphorylation for proteins in Table 3, other proteins show a major p*I* shift upon phosphorylation.[44] Based on our results, all experimental p*I*s for proteins in Table 4 are significantly lower than their nonphosphorylated theoretical p*I* values. Splicing factor, Arg/Ser-rich 9, has a theoretical p*I* of 8.7; however, the detected p*I* for this protein is ~3.7 pH units lower. Analysis of the tryptic digest of the protein using LC−MS/MS revealed three phosphorylations. As shown in Figure 4, G*p*SPHYF*p*SPFRPY, a +2 charged peptide, contains two phosphorylated serines since three product ions created due to the neutral loss of H_3PO_4 were observed. On average, one phosphorylation decreases the p*I* of splicing factor, Arg/Ser-rich 9, by ~1 pH unit. However, when ProMoST was used to calculate the theoretical p*I* for the same protein with three phosphorylations, the calculated value was ~1.5 pH units higher than the measured value. If we plot the calculated change of p*I* upon

(39) Blom, N.; Gammeltoft, S.; Brunak, S. *J. Mol. Biol.* **1999**, *294*, 1351−1362.
(40) Songyang, Z.; Shoelson, S. E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W. G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R. J.; Neel, B. G.; Birge, R. B.; Fajardo, J. E.; Chou, M. M.; Hanafusa, H. *Cell* **1993**, *72*, 767−778.
(41) Songyang, Z.; Blechner, S.; Hoagland, N.; Hoekstra, M. F.; Piwnica-Worms, H.; Cantley, L. C. *Curr. Biol.* **1994**, *4*, 973−982.
(42) Johnson, L.; Lowe, E.; Noble, M.; Owen, D. *FEBS Lett.* **1998**, *430*, 1−11.
(43) Pinna, L. A.; Ruzzene, M. *Biochim. Biophys. Acta* **1996**, *1314*, 191−225.
(44) He, Q.; Skog, S.; Wu, C.; Johansson, A.; Tribukait, B. *Biochim. Biophys. Acta* **1996**, *1289*, 25−30.

Table 4. Proteins with Significant p*I* Shift upon Phosphorylation

protein	acc no.	MW	peptide sequence	phospho (at protein)	obsd phos sites[a]	theor p*I* database[b]	exp p*I*	theor p*I*[c] calcd	pred phos sites[d]
splicing factor, arginine serine-rich 9	Q13242	25 542	ST*p*SYGYSR	S189	3	8.7	5.0−4.8	6.54	Ser 9
			G*p*SPHYF*p*SPFRPY	S211, S216					Thr 1
									Tyr 3
triosephosphate isomerase (Tim) (EC 5.3.1.1)	P60174	26 538	KQ*p*SLGELIGTLNAAK	S21	1	6.4	5.0−4.8	6.14	Ser 3
									Thr 1
									Tyr 1
keratin 5 (*H. sapiens*)	P13647	62 462	TSFTSV*p*SR	S36	5	8.1	4.8−4.6	5.90	Ser 23
			*p*SLYNLGGSK	S64					Thr 1
			ISISTSGG*p*SFR	S82					Tyr 2
			V*p*SLAGACGVGGYGSR	S50					
			*p*SFSTASAITPSVSR	S16					
KIAA1966 protein (*H. sapiens*) (putative splicing factor YT521)	Q96MU7	57 617	GI*p*SPIVFDR	S257	1	5.87	5.4−5.2	6.56	Ser 44
									Thr 2
									Tyr 7
splicing factor, arginine serine-rich 7 (splicing factor 9G8)	Q16629	27 858	*p*SI*p*SLRR	S165, S167	4	11.83	5.4−5.2	12.08	Ser: 41
			*p*SA*p*SPERMD	S231, S233					
hypothetical protein FLJ40432 (*H. sapiens*)	Q8N7R7	34 391	SF*p*SADNFIGIQR	S274	1	6.46	5.2−5.0	6.49	Ser 6
									Thr 1
heterogeneous nuclear ribonucleoprotein A3 (hnRNP A3) (D10S102)	P51991	40 549	SSGpSPYGGGYGSGGGSGGYGSR	S358	1	8.74	5.4−5.2	7.98	Ser 13
									Thr 1

[a−d] These fields are the same as described in Table 3.

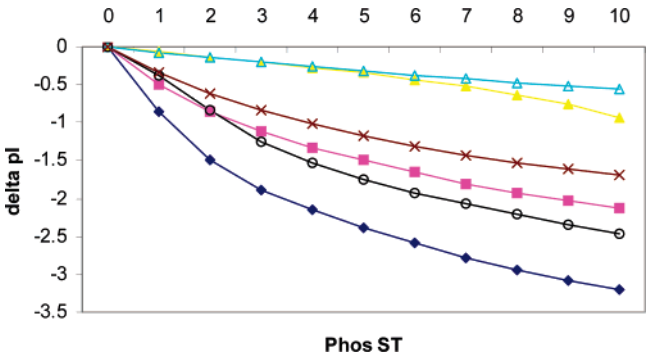

Figure 6. Variation of the theoretical p*I* values as a function of the number of phosphorylations on splicing factor, arginine/serine rich 9 (◆), CK6F (■), splicing factor, arginine/serine rich 7 (▲), KIAA1966 (△), hnRNP A3 (○), and TIM (×). The theoretical p*I* values are calculated using the ProMoST.

phosphorylation as a function of the number of phosphorylations, one can see in Figure 6 that even with 10 phosphorylations the calculated p*I* for splicing factor, Arg/Ser-rich 9, is still ~0.2 pH unit higher than the experimental observation. We estimate from ProMoST that 12 phosphorylations are required to bring the p*I* from 8.7 to ~5.0. Interestingly, 13 very likely phosphorylations are predicted by NetPhos 2.0. The number of phosphorylations predicted using ProMoST and our experimental p*I* matches surprisingly well with that from NetPhos 2.0 although the two methods are based on completely different theories. In the SwissProt database, no phosphorylations have been reported for splicing factor, Arg/Ser-rich 9. It should be pointed out that the theoretical p*I* calculated using ProMoST is 8.42, ~0.3 pH unit lower than the theoretical p*I* in the SwissProt database. Considering the experimental MW value and the results from LC−MS/MS, it is unlikely that the entire p*I* shift is due to such a large number of phosphorylation groups and that other issues involved in calculating the p*I* are involved in this difference.

Another protein in Table 4, triosephosphate isomerase (TIM), also has one detected phosphorylation. The calculated theoretical p*I* for one phosphorylation is 6.14, still much higher than the observed p*I*. Alternatively, five phosphorylations are predicted using NetPhos 2.0. The corresponding calculated p*I* is 5.31 close to the experimental value. Other proteins in Table 4 have one or more phosphorylations detected based upon LC−MS/MS, but the experimental p*I* values are significantly different from the calculated p*I* values. In each case, the number of detected phosphorylations is significantly lower than the predicted number of phosphorylations. In SwissProt, CK5 has one reported phosphoserine, which is detected in the LC−MS/MS analysis—*p*SLYNLGGSK. Four new phosphorylations—TSFTSV*p*SR, ISISTSGG*p*SFR, V*p*SLAGACGVGGYGSR, and *p*SFSTASAITPSVSR—were identified in our analysis. For splicing factor Arg/Ser-rich 7, no phosphorylation has been reported in SwissProt, while four phosphorylations—*p*SI*p*SLRR and *p*SA*p*SPERMD—were identified in our LC−MS/MS analysis. Given that no phosphotyrosine has been identified in our experiment, many phosphorylations on tyrosine might remain undetected in the LC−MS/MS analysis.

The difference in experimental and theoretically calculated p*I* values may be for a number of reasons. First, the pK_a values used in calculation are theoretical pK_a values, while the actual pK_a values

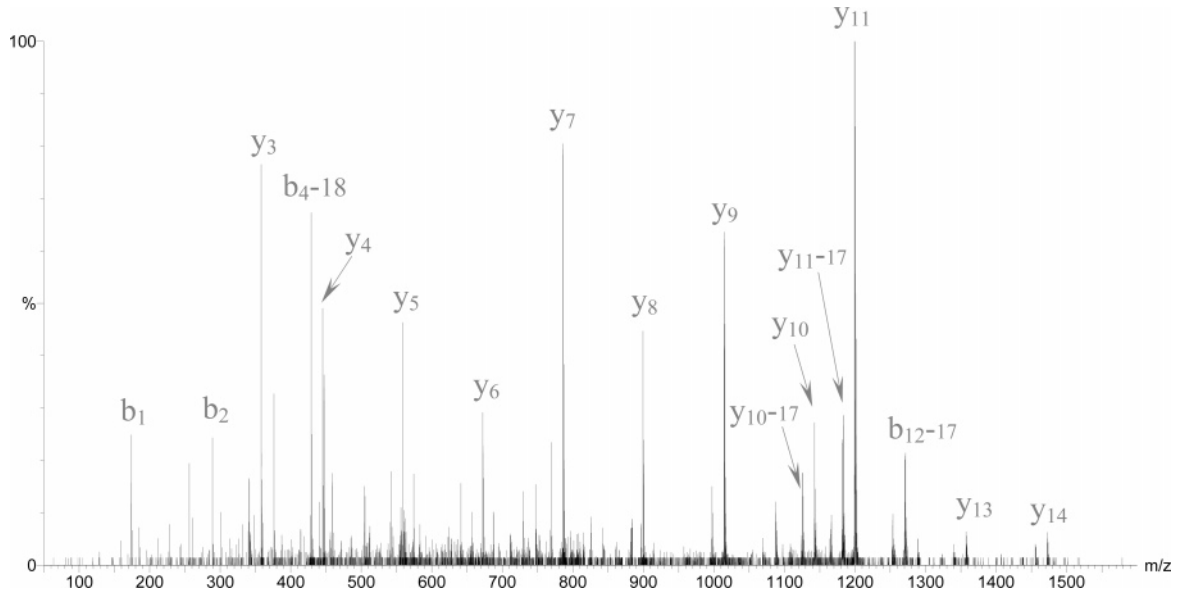

Figure 7. MS/MS spectrum of MDSAGQDINLNSPNK (2+) (N-term acetylated) from tumor protein D52-like 2.

Table 5. Effect of Acetylation on PI

protein	acc no.	theor MW	theor p*I*[a]	acetylation	exp p*I*
tropomyosin (β)	P07951	32 850	4.66	LEKTIDDLEETLASAK + acetyl (K); acetyl (N-term)	4.8−4.6
tumor protein D52-like 2	O43399	22 238	5.26	MDSAGQDINLNSPNK + acetyl (N-term)	5.2−5.0
heterogeneous nuclear ribonucleoprotien F	P52597	45 672	5.38	MLGPEGGEGFVVK + acetyl (N-term)	5.2−5.0
CK7	P08729	51 286	5.5	SIHFSSPVFTSR + acetyl (N-term)	5.4−5.2
hypothetical protein FLJ10292	Q96A72	17 276	5.96	AVASDFYLR + acetyl (N-term)	5.2−5.0
protein disulfide isomerase ER60 precursor	P30101	56 782	5.98	KFLDAGHK + 2acetyl (K), acetyl (N-term)	5.2−5.0

[a] Theoretical p*I* without acetylation(s).

under the experimental conditions might be very different from the theoretical values. In the theoretical model for the p*I* calculation, the interactions between amino acids and phosphate groups were ignored. However, it has been shown in the artificial neural networks analysis that the residues close to the phosphorylation sites do affect the probability of phosphorylation on the site[39]. In addition, many modifications very likely remain unidentified.

An interesting comparison between the proteins in Table 3 and Table 4 becomes apparent when studying the p*I* shift upon phosphorylation. Compared with proteins in Table 4, proteins in Table 3 are more acidic with a p*I* close to pH 5.0; while proteins in Table 4 have p*I*s that vary from pH 8.7 to pH 6.4. These results indicate that basic proteins will result in a significant p*I* shift under phosphorylation while proteins with p*I* close to or lower than pH 5.0 will not exhibit a significant p*I* shift when phosphorylated. Thus, a significant p*I* shift usually suggests posttranslational modifications, such as truncation and phosphorylation; while the lack of a significant p*I* shift does not exclude the presence of posttranslational modifications. In the case of phosphorylation, a p*I* shift is a good indication of phosphorylation for proteins with p*I* > 6.4, but this is not necessarily true for more acidic proteins.

Effect of Acetylation on the p*I* of Proteins. Acetylation at the N-terminal and side chain removes amino group(s) and results in the change of the acid−base balance. A p*I* shift for acidic and neutral proteins due to acetylation has been observed.[4] The p*I* shift is generally small (<0.2 pH) with some exceptions, which may be due to other modifications. In our results, all proteins in Table 5 are acetylated on the N-terminal end, which has been confirmed by MS/MS. The MS/MS spectrum of MDSAGQDINLNSPNK (2+) from D52-like 2 is presented in Figure 6, where all b ions, including b_1, have a 42-Da shift to higher m/z, while the y ions do not show any shift. In Table 5, the theoretical p*I* of tropomyosin β, tumor protein D52-like 2, heterogeneous nuclear ribonucleoprotein F, and CK7 in their nonacetylated form is close to the p*I* of their experimental acetylated counterparts. An additional acetylation on the K of tropomyosin does not significantly affect the p*I*. The detected p*I* of both hypothetical protein FLJ10292 and protein disulfide isomerase is very different from their theoretical nonacetylated form, which may result from other as yet unidentified modifications.

CONCLUSION

It is shown that the p*I* of the intact protein is an important parameter for confirming the identity of a protein, where the theoretical p*I* predicted from the databases often matches the p*I* of unmodified proteins. In combination with the intact MW value, and peptide map, the p*I* aids in confirming protein identification

and often indicates whether the protein is modified. It is shown that truncations and deletions often result in changes in the p*I* related to the number of positively and negatively charged amino acids that have been deleted. Such truncations are also readily observed from the experimental MW value. Phosphorylations often result in a shift of the p*I* to lower values as determined by the number of phosphorylations and the protein involved. The presence of several phosphorylation groups may result in the shift of the p*I* by several pH units. These shifts due to phosphorylation are often observed for proteins of p*I* > 6.4 but may not be observed for proteins where the p*I* value is 5.0 or lower. Acetylation may also cause minor shifts in the p*I*. The p*I* is a parameter that can be associated with modified isoforms of a protein and will be an essential piece of information in the mapping of proteins in human cells.

ACKNOWLEDGMENT

This work was supported in part by the National Institutes of Health under Grant R01 GM 49500 (D.M.L.) and the National Cancer Institute under Grants R21CA83808(D.M.L., F.R.M.) and R01CA90503 (F.R.M., D.M.L.). The MALDI-TOF MS instrument used in this work was funded by the National Science Foundation under Grant DBI 99874.

Received for review October 12, 2004. Accepted February 17, 2005.

AC048494W