

Directions in Discovery

This installment of "Directions in Discovery" looks at affinity-based isolation and enrichment techniques that chemists can use to solve problems in proteomics.

Tim Wehr
Directions in Discovery Editor

Affinity Selection Techniques for Proteomic Studies

Proteomics, the study of the complement of proteins expressed in an organism as a function of its metabolic state, is a key contributor to the drug discovery process. Expression proteomics, which seeks to determine which proteins are produced to greater or lesser degrees in response to a disease state, can be used to elucidate targets for therapeutic intervention. Separation technologies such as gel electrophoresis and liquid chromatography (LC) play central roles in expression proteomics, and earlier installments of "Directions in Discovery" have addressed traditional approaches and newer strategies such as multidimensional LC techniques (1,2).

In a typical comparative proteomics experiment, proteins are isolated from control and experimental populations and subjected to proteolytic cleavage, and their cleavage products are identified by LC coupled with tandem mass spectrometry (LC-MS-MS). Measurements of differential protein expression are achieved by introducing stable isotopes into the experimental sample. This procedure permits the identification of control and experimental proteins in the mass spectra as peak pairs that differ in mass corresponding to the stable isotope. Quantitative information is provided by the relative abundance of the control and experimental peaks in the mass spectrum. Isotopic markers can be introduced during proteolytic cleavage by using trypsin in the presence of H_2^{18}O (3). This approach also can be used to introduce ^{18}O during deglycosylation of N-linked oligosaccharides with PNGase F (4). An alternative approach for isotopic labeling is the acylation of primary amines with ($^2\text{H}_3$)-*N*-acetoxy succinimide (5) or esterification of C-terminal and side-chain carboxyl groups with $^2\text{H}_3$ -methanol (6). A limitation of using deuterium as an isotopic marker is that a labeled peptide will be less hydrophobic and can be resolved partially from the unlabeled peptide during reversed-phase chromatography. Partial

resolution will lead to variations in the $^2\text{H}/^1\text{H}$ ratio across the peak. The deuterium isotope effect will be magnified for smaller peptides, larger labeling compounds, the number of deuterium atoms, and, most important, the location of the deuterium atoms in a hydrophobic portion of the derivatizing agent (7).

A fundamental obstacle to obtaining quantitative information about protein expression is the enormous number of protein species present in a cell. At any given time, the cell may express as many as 20,000 individual proteins, and proteolytic cleavage can generate as many as 100 peptides for each protein. Assuming a peak capacity of 75–100 for reversed-phase gradient elution and unit mass resolution across a mass range of 3000, the resolving capacity of single-dimension LC-MS will be inadequate for analyzing samples that contain more than one million components. Two solutions to this problem are the coupling of two or more chromatographic modes to increase chromatographic resolution and the use of affinity selection techniques. Multidimensional chromatography was discussed previously, so this installment of "Directions in Discovery" will focus on affinity techniques (2). Affinity selection for comparative proteomics has been reviewed recently, and this forms the basis for the overview presented here (8).

A preliminary affinity selection step can accomplish two goals in a proteomics experiment. First, it can reduce the complexity of the peptide digest and permit sufficient resolution in the chromatographic and mass domains to enable quantitative estimates of protein abundance and accurate protein identification. However, the affinity prefractionation must not bias the experiment against proteins that could be of importance in meeting the goals of the study. Therefore, the peptides yielded by the affinity selection technique should represent the entire proteome as much as possible. Second, affinity selection can tar-

get classes of proteins that are involved in particular cellular processes and their regulation. Protein modifications such as phosphorylation and glycosylation are known to be involved in cellular signal transduction and modulating protein function, transport, and lifetime. Affinity capture techniques based upon these modifications can narrow the search for the proteins of interest.

Isotope-Coded Cysteine Affinity Tags

The use of an isotope-coded cysteine affinity tag currently is the most popular affinity method. I have described this technique in an earlier column, so I will review it only briefly (1). The isotope-coded cysteine affinity tag approach uses a reagent with a biotin affinity group, a sulfhydryl-specific iodoacetate moiety, and a linker that carries eight ^1H or ^2H atoms (9). Proteins isolated from control and experimental cells are reacted with light and heavy isotope-coded cysteine affinity tag reagents, respectively, in the presence of a reducing agent. This step specifically tags cysteinyl residues on the proteins. Control and experimental extracts then are mixed and subjected to proteolytic cleavage.

Cysteine-containing peptides are harvested selectively by passage of the mixture through an avidin affinity column, and these peptides are analyzed by capillary reversed-phase LC-MS-MS. Peptides of identical sequences from the control and experimental samples appear as paired masses with differences of 8 amu (singly charged peptides) or 4 amu (doubly charged peptides). Because cysteine occurs in approximately 90% of human proteins but only 17% of the tryptic peptides contain cysteine, the isotope-coded cysteine affinity tag technique meets the goal of reducing the complexity of the mixture approximately fivefold while still representing the majority of the proteome. In addition, the presence of multiple cysteine residues in a protein serves as an internal control in database searches, because these residues will appear as replicate positive identifications for the same protein. The isotope-coded cysteine affinity tag method has three limitations. First, it will miss biologically important proteins if they are among the 10% that contain no cysteine. Second, small peptides or those that contain multiple cysteine residues might exhibit partial or complete resolution of the heavy and light derivatives. Third, the elu-

tion of peptides from the avidin affinity column might not be quantitative.

Histidine Affinity Selection

Immobilized metal affinity chromatography can be used to capture peptides that can form coordination complexes with metals bound to the immobilized metal affinity chromatography ligand (10,11). For example, an immobilized metal affinity chromatography column carrying iminodiacetate functions complexed with copper can selectively bind peptides that contain histidine, cysteine, tryptophan, and N-terminal amino groups. The procedures used to prepare protein samples for enzymatic digestion can increase the selectivity of immobilized metal affinity chromatography for histidine-containing peptides. Specifically, the reduction and alkylation steps used before tryptic digestion eliminate cysteine binding to the column. If differential isotopic labeling by acylation is used to detect up- or down-regulation of proteins, the introduction of acyl groups blocks the binding of N-terminal amino groups and tryptophan to the immobilized metal affinity chromatography support (12). The steps in immobilized metal affinity chromatography include loading the column with the metal solution, introducing the sample, washing unbound peptides from the column with a buffered salt solution, and eluting the bound peptides with a competing species such as imidazole. Peptide release can be accomplished by step or gradient elution.

The frequency of histidine in the human proteome is similar to that of cysteine. It is found in 83% of proteins and in 15% of the peptides generated by proteolysis of the proteome (8). Therefore, as with cysteine selection in the isotope-coded cysteine affinity tag technique, histidine selection in immobilized metal affinity chromatography provides a fourfold reduction in sample complexity with little loss in the representation of the complete proteome in the analysis. Recently, Wang and co-workers (13) reported the double selection of peptides using cysteine and histidine residues for proteomic studies. This selection was accomplished by first selecting cysteine-containing peptides by covalent chromatography using thiol disulfide exchange. After the release of cysteine-containing peptides, histidine-containing peptides were captured by passage through a copper-loaded immobilized metal affinity chromatography column. This double-selection technique

Table I: Commercially available lectins

Lectin	Source	Affinity
AAA	<i>Anguilla anguilla</i>	α -L-Fucosyl residues
ABA	<i>Agaricus bisporus</i>	Carbohydrate receptors
BPA	<i>Bauhinia purpurea</i>	N-Acetyl-D-galactosamine and D-galactose
BSI-A4	<i>Bandeiraea simplicifolia</i>	Terminal N-acetyl- α -D-galactosyl residues
BSI-B4	<i>Bandeiraea simplicifolia</i>	Terminal α -D-galactosyl residues
CCA	<i>Caragana arborescens</i>	N-Acetyl-D-galactosamine
Codium fragile	<i>Codium fragile</i>	N-Acetyl-D-galactosamine
Concanavalin A	<i>Canavalia ensiformis</i>	Terminal α -D-mannosyl and α -D-glucosyl residues
ECoRA	<i>Erythrina corallodendrom</i>	N-Acetylglucosamine, N-acetyl-D-galactosamine, lactose, D-galactose
EEA	<i>Eunymus europaeus</i>	α -D-Gal(1 > 3)[α -L-Fuc(1 > 2)]- β -D-Gal(1 > 3/4)- β -D-GlcNac
GNL	<i>Galanthus nivalis</i>	Terminal α -D-mannosyl residues of glycoconjugates
HAA	<i>Helix aspersa</i>	Terminal N-acetyl- α -D-galactosaminyl residues
HPA	<i>Helix pomatia</i>	Terminal N-acetyl- α -D-galactosaminyl residues
LcH	<i>Lens culinaris</i>	Terminal α -D-mannosyl and α -D-glucosyl residues
LEA	<i>Lycopersicon esculentum</i>	N-Acetyl- β -D-glucosamine oligomers
Limulus polyphemus	<i>Limulus polyphemus</i>	N-Acetylated D-hexosamines
MMA	<i>Maackia amurensis</i>	Sialic acid-containing glycoconjugates
MPA	<i>Maclura pomifera</i>	α -D-Galactosyl and N-acetyl-D-galactosaminyl residues
NPA	<i>Narcissus pseudonarcissus</i>	Terminal and internal α -D-mannosyl residues of glycoconjugates
PSA	<i>Pisum sativum</i>	Terminal α -D-mannosyl residues
Psophocarus tetragonolobus	<i>Psophocarus tetragonolobus</i>	N-Acetyl-D-galactosamine
PWM	<i>Phytolacca americana</i>	N-Acetyl- β -D-glucosamine oligomers
RCA ₆₀	<i>Ricinus communis</i>	N-Acetyl-D-galactosamine
RCA ₁₂₀	<i>Ricinus communis</i>	Terminal β -D-galactosyl residues
SBA	<i>Glycine max</i>	N-Acetyl-D-galactosamine
SNA	<i>Sambucus nigra</i>	α -NeuNAC-[2 > 6]-Gal; α -NeuNAC [2 > 6]-GalNac; α -NeuNAC[2 > 3]-Gal
STA	<i>Solanum tuberosum</i>	N-Acetyl- β -D-glucosamine oligomers
Tetragonolobus purpureas	<i>Tetragonolobus purpureas</i>	α -L-Fucosyl residues
UEA	<i>Ulex europaeus</i>	L-Fucose, N,N'-diacetylchitobiose
VAA	<i>Viscum album</i>	Terminal β -D-galactosyl residues
VFA	<i>Vicia faba</i>	D-Mannose, D-glucose
VVA	<i>Vicia villosa</i>	N-Acetyl-D-galactosamine
WFA	<i>Wisteria floribunda</i>	N-Acetyl-D-galactosamine
WGA	<i>Triticum vulgare</i>	N-Acetyl- β -D-glucosaminyl residues, N-acetyl- β -D-glucosamine oligomers

reduced sample complexity by 95% but did not seriously compromise the coverage of the proteome (76% of human proteins have at least one peptide that contains both cysteine and histidine residues).

Glycosylation Affinity Selection

Posttranslational modification of proteins caused by the addition of carbohydrate groups to certain amino acids can affect protein function and can be associated with the metabolic state of the cell or with disease progression. Affinity methods for selecting glycosylated proteins and peptides therefore is of interest in proteomic studies. Proteins can be glycosylated by the addition of saccharides at one or more specific sites in their primary sequences, and the structure of the carbohydrate moiety at a particular site can be heterogeneous. Oligosaccharides can be assembled stepwise at serine and threonine residues to form O-linked glycosylation sites or can be pre-assembled and attached at asparagine residues to form N-linked glycosylation sites. The resulting glycosylation patterns

can be quite complex, with a single polypeptide amino acid sequence represented in a sample by multiple glycoforms. In terms of the separation problem, a single sequence in a protein or deoxyribonucleic acid (DNA) database can account for a dozen or more components in a two-dimensional gel or a multidimensional LC experiment.

Lectins provide a powerful approach for selectively isolating glycoproteins and glycopeptides from a complex mixture (10,14,15). Lectins are plant proteins that bind to carbohydrate groups, and they differ in their binding specificity. Concanavalin A is useful for selecting polypeptides with N-linked oligosaccharides. Concanavalin A has a high affinity for N-linked hybrid and high-mannose oligosaccharides, less affinity for complex diantennary oligosaccharides, and little affinity for complex N-linked tri- and tetraantennary oligosaccharides. The lectin from *Bandeiraea simplicifolia* has a narrow selectivity and binds specifically to O-linked N-acetylglucosamine. Some of the

commercially available lectins are listed in Table I.

Historically, lectins have been used in glycobiology studies to isolate glycoproteins before analyzing the carbohydrate moieties. In proteomic studies, the masses and sequences of peptides derived from a protein are used for database searches to identify and quantify expressed proteins. Therefore, lectins are used in proteomics studies simply to isolate this class of post-translationally modified proteins. The carbohydrate group can be retained in the analysis in the case of homogeneous carbohydrate structures, and the oligosaccharide mass can be subtracted from the observed peptide mass for database searches. In the case of heterogeneous carbohydrate structures, the oligosaccharides can be enzymatically cleaved from the polypeptides before mass and sequence analysis.

Figure 1 presents a schematic for glycoprotein analysis in a proteomic study (15). After reduction, alkylation, and proteolysis, the peptides are loaded onto a lectin column, and nonglycosylated

peptides are washed from the column. Glycopeptides then are eluted from the column with an appropriate displacing agent such as methyl- α -D-mannopyranoside for a concanavalin A column or *N*-acetylglucosamine for a *B. simplicifolia* column (15).

For glycopeptides that contain homogeneous carbohydrate moieties, the eluate from the lectin column can be subjected to reversed-phase LC–MS or LC–MS–MS analysis directly. For glycopeptides with heterogeneous carbohydrate structure, the variations in the oligosaccharide portion of the peptide can change the chromatographic behavior of the glycopeptide and result in reversed-phase LC resolution of a single polypeptide into multiple peaks. The resultant chromatogram will increase in complexity, and, in MS analysis, the peptide will be distributed across several

masses depending upon the carbohydrate heterogeneity.

Fragmentation of the oligosaccharide in the mass spectrometer also can confound spectral interpretation. In this situation, enzymatic cleavage of carbohydrate groups from the glycopeptide before reversed-phase LC–MS simplifies the analysis. Digestion with the enzyme PNGase F can be used to cleave N-linked oligosaccharides attached to the protein at asparagine residues. PNGase cleavage leaves an aspartate residue at the cleavage site.

Affinity Selection Based upon Phosphate Groups

Phosphorylation is one of the most frequent and important posttranslational modifications of proteins. Approximately one-third of all mammalian cellular proteins are phosphorylated. Phosphorylation

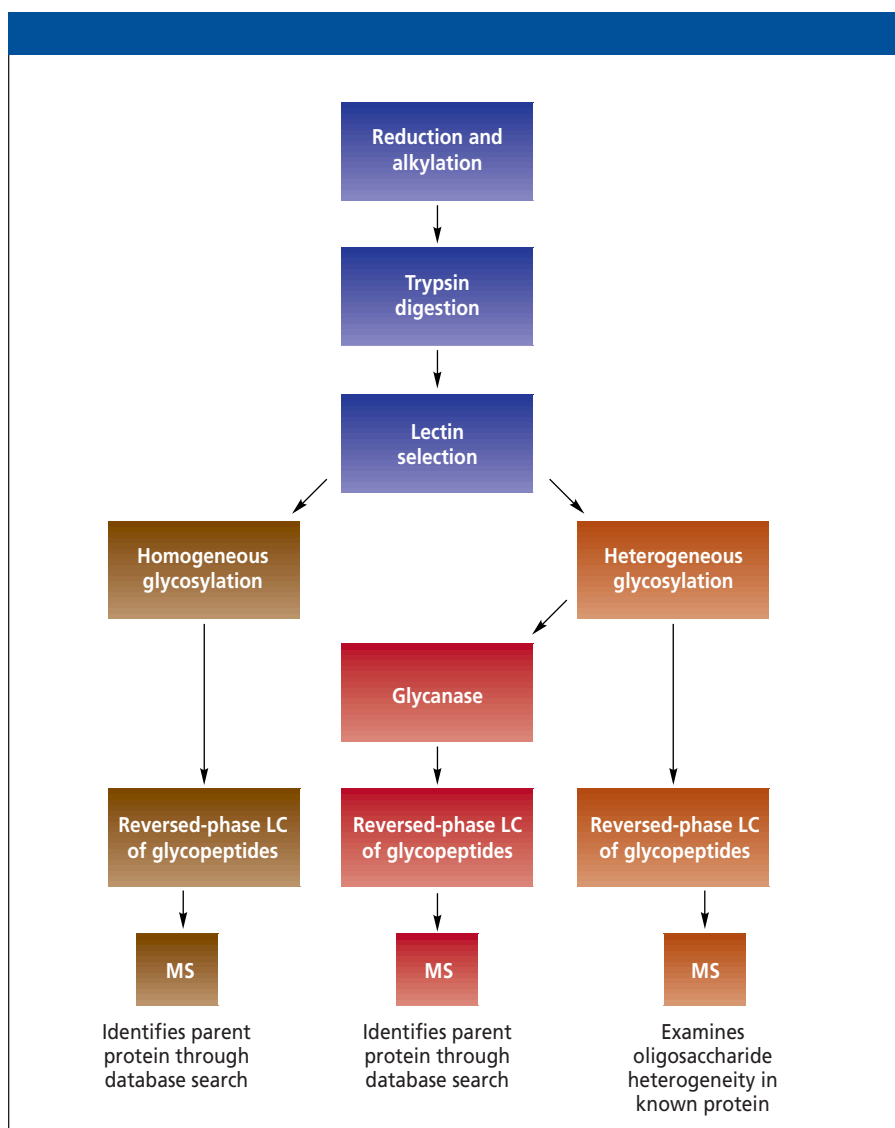


Figure 1: Protocol for affinity selection of glycopeptides. (Reproduced from reference 15 with permission from Elsevier Science BV.)

of serine accounts for roughly 90% of these modifications, with phosphorylation of threonine and tyrosine residues accounting for 10% and 0.1% of the total, respectively. Reversible phosphorylation is known to be involved in the regulation of cellular events and processes, including signal transduction, gene expression, metabolism, and cell growth, division, and differentiation. Knowledge of the sites of protein phosphorylation and the correlation of phosphate residency with metabolic changes will be important for analysts' understanding of cellular regulation. An obstacle in determining the state of protein phosphorylation is the huge complexity of the proteome and the occurrence of potentially important phosphoproteins at low concentrations in the cell. Affinity enrichment of phosphoproteins serves to eliminate interferences and to increase signal in protein identification experiments, which is particularly important in the case of polypeptides phosphorylated at multiple sites because they exhibit signal suppression in sequencing experiments using positive-ion electrospray ionization MS-MS.

Analysts have used two affinity approaches for phosphoprotein enrichment. The first is chromatography on immobilized metal affinity chromatography columns complexed with Fe^{3+} , Ga^{3+} , or other metals (16–20). Protein samples are reduced, alkylated, and then cleaved with a proteolytic enzyme such as trypsin. Peptide cleavage products are loaded onto

the immobilized metal affinity chromatography column under acidic conditions (pH 2.5–3.5), unbound nonphosphopeptides are removed from the column with an acidic wash solution, and phosphopeptides are eluted under alkaline conditions (\sim pH 10). Posewitz and Tempst (17) evaluated a variety of metals, including Fe^{3+} , Ga^{3+} , Al^{3+} , and Zr^{3+} , for immobilized metal affinity chromatography enrichment of phosphopeptides. They observed the best selectivity and recovery with iminodiacetate columns complexed with Ga^{3+} . They found increased binding of nonphosphopeptides when loading was performed outside the pH 2.0–3.5 range.

Stensballe and colleagues (19) reported that excessive loading of peptide samples increased the nonspecific binding of nonphosphopeptides (19). In another study, Zhou and co-workers (18) evaluated Fe^{3+} - and Ga^{3+} -complexed nitrilotriacetate-immobilized metal affinity chromatography materials for direct ionization of phosphopeptides in matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) MS. The two metals proved to be complementary in their performance. The Ga^{3+} -immobilized metal affinity chromatography material was preferable for isolation of peptides with multiple phosphate groups, and the Fe^{3+} -immobilized metal affinity chromatography complexes were preferable for monophosphorylated peptides because multiphosphorylated peptides were

strongly bound. In MS analysis of peptides eluted from immobilized metal affinity chromatography columns, the presence of phosphate groups can be confirmed by treating the sample with alkaline phosphatase and reanalyzing the treated sample. Phosphopeptides will exhibit a loss of 80 amu for each phosphate group removed by enzyme treatment, and the mass of nonphosphopeptides will be unchanged.

The contamination of immobilized metal affinity chromatography-separated phosphopeptides with nonphosphorylated peptides complicates mass and sequence analysis by MALDI-TOF and electrospray ionization MS-MS. Spectra tend to be more complex, and confirmation of phosphate presence by enzyme treatment adds an additional step. Another approach for affinity enrichment of phosphopeptides uses site-specific chemical modification of the phosphoamino acids followed by the introduction of a highly specific affinity ligand (21,22). Under strongly alkaline conditions, phosphate groups on serine residues will undergo β -elimination to form a dehydroalanyl residue, and threonine yields dehydroamino-2-butyric acid. Both of these elimination products react readily with ethanedithiol. The remaining free thiol group of ethanedithiol can be linked to a biotin affinity tag for affinity isolation on an avidin column (Figure 2) (21).

The advantage of this approach is the nearly quantitative recovery of biotin-tagged phosphopeptides from a peptide mixture using immobilized monovalent avidin. In addition, the identity of phosphopeptides is confirmed by the presence of the 446-amu biotin fragment in MS-MS experiments and the appearance of the $M + H$ precursor ion as a $\Delta 18$ -amu doublet in the mass spectrum, which represents the open and closed forms of the biotin tag. This technique has several limitations. First, O-glycosylated amino acids undergo the same reaction and yield dehydroalanyl and dehydroamino-2-butyric acid residues. This reaction can be minimized by preferentially enhancing the rate of phosphate elimination by performing the reaction in barium hydroxide rather than sodium hydroxide. Second, high concentrations of ethanedithiol are necessary to suppress protein degradation under alkaline conditions. Third, the pretreatment of the sample with peroxide is necessary to block reactive cysteinyl residues. However, peroxide treatment converts methionine residues to methionine sul-

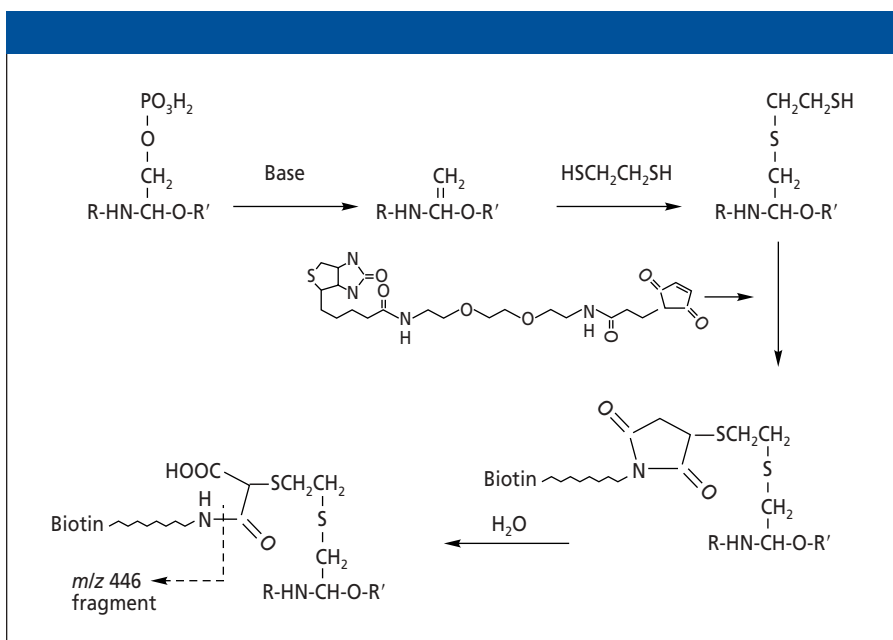


Figure 2: Scheme for chemical conversion of a phosphoserine residue to a biotinylated residue. The dashed line represents the cleavage site that produces a 446-amu biotin fragment in MS-MS experiments. (Reproduced from reference 21 with permission from the Nature Publishing Group.)

phones and oxidizes tryptophan residues. Finally, racemization during ethanedithiol addition produces diastereomers that can be resolved chromatographically (22).

The phosphopeptide biotinylation procedure has been used as the basis for introduction of a phosphoprotein isotope-coded affinity tag (23). In this approach, Goshe and colleagues labeled the ethanedithiol reagent with either four hydrogen or four deuterium atoms. The free sulfhydryl of the ethanedithiol addition product was tagged with iodoacetyl-PEO-biotin for affinity isolation. After tryptic digestion, the affinity purification of tagged phosphopeptides and LC-MS analysis were performed in a manner similar to previous studies. In this case, isotope coding of the phosphopeptides can permit comparative measurement of phosphorylation in two distinct protein samples.

Summary

Analysts can use affinity-based isolation and enrichment techniques to solve two analytical problems in proteomics. First, they can narrow the complexity of the sample to facilitate determination of peptide mass and sequence for accurate protein identification. For this, it is necessary that the affinity group be present in a small subset of the peptides but also that this subset represent as large a fraction of the proteome as possible. Cysteine and histidine affinity selection are two examples of this approach. Incorporating isotope-coded tags into an affinity ligand enables the quantitative measurement of differential protein expression. Second, they can select peptides carrying posttranslational modifications that are functionally important in cellular processes and regulation. Glycosylation and phosphorylation are two important modifications that can be used as the basis for affinity enrichment.

References

- (1) T. Wehr, *LCGC* **19**(7), 702–711 (2001).
- (2) T. Wehr, *LCGC* **20**(10), 954–962 (2002).
- (3) X. Yao, A. Freas, J. Ramirez, P.A. Demirev, and C. Fenselau, *Anal. Chem.* **73**, 2836–2842 (2001).
- (4) B. Kuster and M. Mann, *Anal. Chem.* **71**, 1431–1440 (1999).
- (5) J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, and F.E. Regnier, *J. Chromatogr. B* **745**, 197–210 (2000).
- (6) D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, and E. Kolker, *Rapid Commun. Mass Spectrom.* **15**, 1214 (2001).
- (7) R. Zhang, C.S. Sioma, R.A. Thompson, L. Xiong, and F.E. Regnier, *Anal. Chem.* **74**, 3662–3669 (2002).
- (8) F.E. Regnier, L. Riggs, R. Zhang, L. Xiong, P. Liu, A. Chakraborty, E. Seeley, C. Sioma, and R.A. Thompson, *J. Mass Spectrom.* **37**, 133–145 (2002).
- (9) S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold, *Nat. Biotechnol.* **17**, 994–999 (1999).
- (10) J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, and F.E. Regnier, *J. Chromatogr.* **745**, 197–210 (2000).
- (11) A. Amini, A. Chakraborty, and F.E. Regnier, *J. Chromatogr. B* **772**, 35–44 (2002).
- (12) P. Hansen, G. Lindeberg, and L. Andersson, *J. Chromatogr.* **627**, 125–135 (1992).
- (13) S. Wang, X. Zhang, and F.E. Regnier, *J. Chromatogr. A* **949**, 153–162 (2002).
- (14) M. Geng, J. Ji, and F.E. Regnier, *J. Chromatogr. A* **870**, 295–313 (2000).
- (15) M. Geng, X. Zhang, M. Bina, and F.E. Regnier, *J. Chromatogr. B* **752**, 293–306 (2001).
- (16) L.M. Nuwaysir and J.T. Stults, *J. Am. Soc. Mass Spectrom.* **4**, 662–669 (1993).
- (17) M.C. Posewitz and P. Tempst, *Anal. Chem.* **71**, 2883–2892 (1999).
- (18) W. Zhou, B.A. Merrick, M.G. Khaledi, and K.B. Tomer, *J. Am. Soc. Mass Spectrom.* **11**, 273–282 (2000).
- (19) A. Stensballe, S. Andersen, O.N. Jensen, *Proteomics* **1**, 207–222 (2001).
- (20) A. Schlosser, J. Bodem, D. Bossemeyer, I. Grummt, W.D. Lehmann, *Proteomics* **2**, 911–918 (2002).
- (21) Y. Oda, T. Nagasu, and B.T. Chait, *Nat. Biotechnol.* **19**, 379–382 (2001).
- (22) M. Adamczyk, J.C. Gebler, and J. Wu, *Rapid Commun. Mass Spectrom.* **15**, 1481–1488 (2001).
- (23) M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstras, and R.D. Smith, *Anal. Chem.* **73**, 2578–2586 (2001).

Tim Wehr

"Directions in Discovery" editor Tim Wehr is associate director of training at Rheodyne LLC, the LC Resources Training Group, Walnut Creek, California. Direct correspondence about this column to



"Directions in Discovery," LCGC, 859 Willamette Street, Eugene, OR 97401, e-mail lcgcedit@lcgcmag.com.