TargetDiscovery
FROM OMICS TO KNOWMICS

4015 Fabian Way, Palo Alto, CA 94303
Tel 650-812-8100 | Fax 650-812-8130
www.targetdiscovery.com

IDBEST™ FOR PROTEIN ISOFORM DISCOVERY AND CHARACTERIZATION

Mass spectrometric (MS) proteomic methods provide the ability to identify proteins from the primary sequence that is not accessible by other proteomic methods (e.g., gels and immunoassays). Global MS proteomic profiling methods—such as multidimensional protein identification (MudPIT) and solid phase or retentate chromatography (Ciphergen ProteinChip®)—involve separating, quantifying, and identifying as many differentially-expressed proteins as possible from a tissue sample. These open methods have proved useful for some biomarker discovery applications, where the research objective is to find hitherto unknown proteins involved in a disease state[1] or toxicological reaction.[2]

Growing Importance of Protein Isoforms as Biomarkers

Current immunoassay (IA) diagnostic methods are often limited to quantification of the up- or down-regulation of the total protein (sum of all isoforms). Current breast cancer biomarkers, such as kallikrein 3 (PSA)[3] and basic fibroblast growth factor (bFGF)[4] alone and in combination,[3] are not 100% diagnostic or prognostic, still leading to both false positive and false negative diagnoses. Any one biomarker may not be completely reliable for the diagnosis or prognosis of breast cancer because immunoassays (IA) may miss important molecular-level information (e.g., mutations, splice variants, and post-translational modifications [PTMs]) that does not affect the epitopes targeted by mAb. For example, Peracaula and coworkers[5] recently showed that the relative abundance of a specific isoform of PSA may be more diagnostic of prostate cancer than the current FDA-approved total and free PSA ratio assay. In another example, it has been shown that the surface marker CD44 plays a role in cell-matrix interactions in breast cancer.[6] Of the breast carcinomas studied, 44% were positive for CD44 by reaction with an mAb; however, only one isoform of this marker, CD44v3, was significantly correlated with the presence of metastases to lymph nodes.[6] Therefore, unless an mAb is raised against every mutation, splice variation, or PTM, many of which are currently unknown, it would be impossible to detect all molecular-level protein variations by IA alone. MS detection uniquely allows post-translational modifications and mutations (often unnoticed in immunoassays) to be detected, increasing the clinical relevance of the biomarker.

---

[1] Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A., "Use of proteomic patterns in serum to identify ovarian cancer," Lancet; 2002; 359;572-7.

[2] Steiner, S. and Anderson, N.L., "Expression profiling in toxicology—potentials and limitations," Toxicol Lett; 2000; 112-113:467-71.

[3] Sauter ER, Klein G, Wagner-Mann C, and Diamandis, EP. Prostate-specific antigen expression in nipple aspirate fluid is associated with advanced breast cancer. Cancer Detect. Prev. 2004;28: 27-31.

[4] Hsuing R, Zhu W, Klein G, Qin W, Rosenberg A, Park P, Rosato E, and Sauter ER. High basic fibroblast growth factor levels in nipple aspirate fluid are correlated with breast cancer. Cancer J. 2002; 8:303-310.

[5] Peracaula, R., Tabares, G., Royle, L., Harvey, D.J., Dwek, R.A., Rudd, P.M., de Llorens, R., "Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins," Glycobiology; 2003; 13:457-470.

[6] Rys J, Kruczak A, Lackowska B, Jaszcz-Gruchala A, Brandys A, Stelmach A, Reinfuss M. The role of CD44v3 expression in female breast carcinomas. Pol. J. Pathol. 2003;54: 243-47.

TargetDiscovery
FROM OMICS TO KNOWMICS

Enrichment of Low Abundance Biomarkers

However, a major issue in global proteomic methods is the limited dynamic range of the detection methods (Figure 1). Proteins exhibit a very wide dynamic range in concentrations, from 10^5 in bacterial cells,[7] to between 10^7 and 10^8 in human cells,[8] to at least 10^{12} in plasma.[9] Since there is no technique to amplify low abundance proteins, comparable to polymerase chain reaction (PCR) for nucleic acids, both gel and MS global profiling methods often fail to detect clinically-relevant low abundance proteins.[10] The dynamic range in MS is only about 10^2-10^3 between the highest abundance peak and the limit of detection. When coupled with other pre-fractionation methods (e.g., LC or high abunance protein depletion methods) global profiling methods have barely reached the upper range for known clinically-validated biomarkers (Figure 1). For example, MudPIT global proteomic characterization of nipple aspirate fluid (NAF) has been reported for the healthy breast,[11] notably failing to reveal kallikrein 3 (PSA) in healthy NAF. Yet, immunoassay (IA) studies of NAF show that PSA is in significant concentration in the healthy breast and down-regulation of PSA levels are highly correlative to breast cancer.[12]

This dynamic range issue can only be addressed through the application of targeted proteomic (closed) methods, such as immunoassays (IAs). Because peptides resulting from the antibody cannot be distinquished from peptides arising from the target protein in the mass spectrum, it has hitherto been extremely difficult to couple to mass spectrometric detection methods to IA (or other baiting strategies) for enrichment of lower abundance biomarkers. Affinity enrichment-MS approaches have been successfully applied at the intact protein level (top-down proteomics).[13-17] but have not been quantitative unless coupled to an additional assay

---

[7] Ingraham, J.L., Maaløe, O., and Neidhardt, F.C., Growth of the Bacterial Cell, (Sinauer Associates, Sunderland, MA, 1983).

[8] Anderson, N.L. and Anderson, N.G., “Proteome and proteomics: new technologies, new concepts, and new words “;Electrophoresis; 1998; 19:1853-61.

[9] Corthais, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.-C., “The dynamic range of protein expression: a challenge for proteomic research,” Electrophoresis; 2000; 21:1104-15.

[10] Kenyon, G.L., DeMarini, D.M., Fuchs, E., Galas, D.J., Kirsch, J.F., Leyh, T.S., Moos, W.H., Petsko, G.A., Ringe, D., Rubin, G.M., Sheahan, L.C., “Defining the mandate of proteomics in the post-genomics era,” Mol. Cell. Proteomics; 2002; 1(10):763-80.

[11] Varnum, S.M., C.C. Covington, R.L. Woodbury, K. Petritis, L.J. Kangas, M.S. Abdullah, J.G. Pounds, R.D. Smith, and R.C. Zangar. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res. Treat.; 2003; 80:87-97.

[12] Sauter, E.R., M. Daly, K. Linahan, H. Ehya, P.F. Engstrom, G. Bonney, E. Ross, H. Yu, and E. Diamandis. Prostate specific antigen levels in nipple aspirate fluid correlate with breast cancer risk. Cancer Epidemiol. Biomarkers Prev.;1996; 5:967-970.

[13] Nedelkov D, Tubbs KA, Niederkofler EE, Kiernan UA, Nelson RW, “High-throughput comprehensive analysis of human plasma proteins: a step toward population proteomics,” Anal Chem.; 2004; 76(6):1733-7.

[14] Nelson RW, Krone JR, Bieber AL, Williams P., “Mass spectrometric immunoassay,” Anal Chem.; 1995; 67(7):1153-8.

[15] Kiernan UA, Nedelkov D, Tubbs KA, Niederkofler EE, Nelson RW, “Proteomic characterization of novel serum amyloid P component variants from human plasma and urine,” Proteomics; 2004; 4(6):1825-9.

[16] Nedelkov D, Nelson RW, “Detection of Staphylococcal enterotoxin B via biomolecular interaction analysis mass spectrometry,” Appl Environ Microbiol.; 2003; 69(9):5212-5.

[17] Kiernan UA, Tubbs KA, Nedelkov D, Niederkofler EE, Nelson RW, “Detection of novel truncated forms of human serum amyloid A protein in human plasma,” FEBS Lett.; 2003; 537(1-3):166-70.

assay technique (e.g., surface plasmon resonance, SPR).[18-19] However, by using mass defect tags (Box A) TDI has finally enabled affinity enrichment-MS assays. The mass defect element shifts the position of tagged peptide peaks at least –0.1 amu in the mass spectrum from the position of any untagged peptides, allowing peptides arising from the target protein to be readily discriminated.

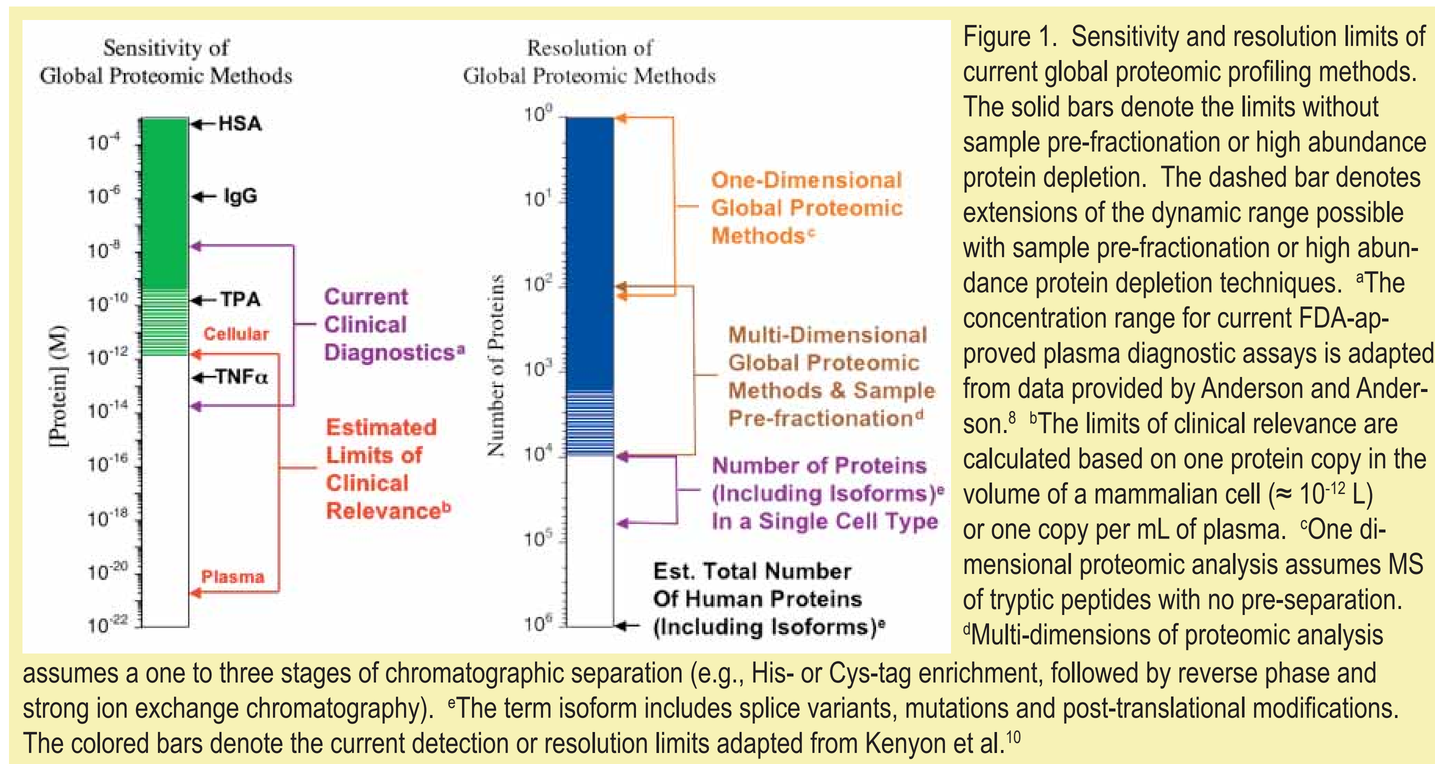

Figure 1. Sensitivity and resolution limits of current global proteomic profiling methods. The solid bars denote the limits without sample pre-fractionation or high abundance protein depletion. The dashed bar denotes extensions of the dynamic range possible with sample pre-fractionation or high abundance protein depletion techniques. [a]The concentration range for current FDA-approved plasma diagnostic assays is adapted from data provided by Anderson and Anderson.[8] [b]The limits of clinical relevance are calculated based on one protein copy in the volume of a mammalian cell (≈ 10^{-12} L) or one copy per mL of plasma. [c]One dimensional proteomic analysis assumes MS of tryptic peptides with no pre-separation. [d]Multi-dimensions of proteomic analysis assumes a one to three stages of chromatographic separation (e.g., His- or Cys-tag enrichment, followed by reverse phase and strong ion exchange chromatography). [e]The term isoform includes splice variants, mutations and post-translational modifications. The colored bars denote the current detection or resolution limits adapted from Kenyon et al.[10]

High Precision Quantification with Stable Isotope Ratio MS

Proteomic methods that do not utilize internal standards, however, often lack the desired precision because of protein recovery differences during sample preparation, this is magnified in MS methods because of differences in ionization efficiency between samples. Difference gel electrophoresis (DIGE), marketed by Amersham Biosciences (a division of GE Healthcare), addresses the recovery variability problem by using two different fluorescent tags, one for the control (or internal standard) and one for the diseased tissue sample, which are mixed and co-separated in a 2-D gel. Stable isotope tagging methods (Box B) are the mass spectrometric equivalent of DIGE.

[18] Nedelkov D, Nelson RW, "Surface plasmon resonance mass spectrometry: recent progress and outlooks," Trends Biotechnol.; 2003; 21(7):301-5.

[19] Nedelkov D, Nelson RW, Kiernan UA, Niederkofler EE, Tubbs KA, "Detection of bound and free IGF-1 and IGF-2 in human plasma via biomolecular interaction analysis mass spectrometry," FEBS Lett.; 2003; 536(1-3):130-4.

[20] Minden, J. and Waggoner, A., "Difference gel electrophoresis using matched multiple dyes," US 6043025 (Mar. 28, 2000).

Box A: The Mass Defect

While an element is defined by the number of nucleons (protons and neutrons) contained in its nucleus, when these nucleons come together to form an atom some energy is liberated due to the efficiency of packing these nucleons together. The amount of energy liberated depends on the number of nucleons packed together into the nucleus and differs for every element and isotope of an element in the periodic table. From Einstein's theory of relativity, this nuclear binding energy has a mass equivalent. Therefore, each element of the periodic table has an actual mass that differs slightly from its mass expected based on the number of nucleons that comprise its nucleus (Figure 2). By convention, this mass defect is defined as zero for ^{12}C and the mass defect of all other elements are scaled to this standard.

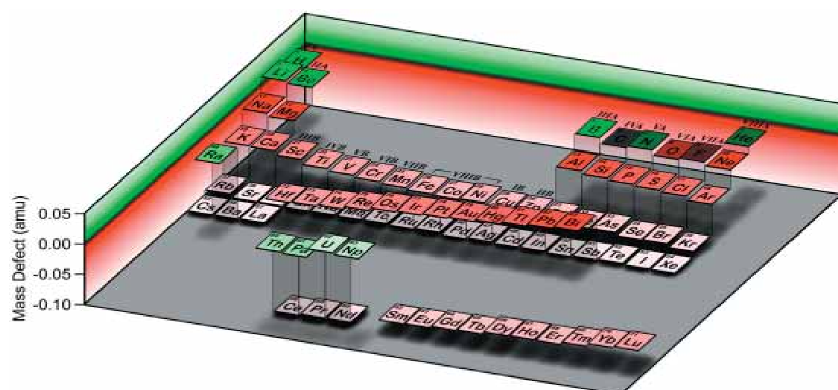

Figure 4. The periodic table of the elements presented in mass defect relief. The mass defect of the most abundant stable isotope of each element is shown based on the [^{12}C] = 0 mass defect standard.

Elements commonly found in biomolecules (e.g., C, H, N, and O) have negligible mass defects. However, the mass defect of elements with atomic numbers between 35 (Br) and 63 (Eu) differs by almost –0.1 amu from that of ^{12}C, which is easily resolvable in today's high resolution time-of-flight (TOF) and Fourier Transform (FT) mass spectrometers. When a mass defect element (e.g., Br) is used to tag a protein, the resulting protein or tagged peptides resulting from the protein are effectively shifted by –0.1 amu from any untagged proteins or peptides in the mass spectrum.[1] With sufficient mass spectrometric resolution, mass defect tagged peptides can be quantitatively deconvolved from non-tagged peptides. This principle forms the basis of IMLS,[2-5] a top-down sequencing method. When stable isotopes are used in conjunction with a mass defect element in the tag this forms the basis of isotope-differentiated binding energy shift tags (ID-BEST).[1, 5] Most of the transition and lanthanide series metals also impart a mass defect to tags that contain them as chelates, such as: element-coded affinity tags (ECAT),[6] and metal-coded affinity tags (MeCAT).[7] The effectiveness of mass defect tags has been independently corrobarated by several groups.[6-9]

---

[1] Schneider, L.V., Hall, M.P., and Brasseur, M.M., "Mass-defect tagging for proteomic analysis," Genetic Eng. News; 2004; 24:28-30.

[2] Schneider LV, Hall MP, Peterson JN. Methods for sequencing proteins. US6379971 (Apr. 30, 2002).

[3] Schneider LV, Hall MP, Peterson JN. Methods for sequencing proteins. US6706529 (Mar. 16, 2004).

[4] Schneider LV, Hall MP, Peterson JN. Methods for sequencing proteins. US6716636 (Apr. 6, 2004).

[5] Schneider, L.V., Hall, M.P., and Petesch, R., "Mass defect labeling for the determination of oligomer sequences," US20020172961 (Nov. 21, 2002).

[6] Whetstone, P.A., Butlin, N.G., Corneille, T.M., and Meares, C.F., "Element-coded affinity tags for peptides and proteins," Bioconj. Chem.; 2004; 16:3-6.

[7] Krause, M., Scheler, C., Böttger, U., Weisshoff, H., Linscheid, M., "Method and reagent for specifically identifying and quantifying on or more proteins in a sample," WO04001420 (Dec. 31, 2003).

[8] Hernandez, H., Vichchulada, P., Niehauser, S., and Amster, I.J., "Improvements in mass defect labeling for shotgun proteomic analysis," Poster presented at the 52nd Ann. Mtg., Amer. Soc. Mass Spectrom., Nashville, TN (May 23-26, 2004).

[9] Hall, M.P., Ashrafi, S., Obegi, I., Petesch, R., Peterson, J.N., and Schneider, L.V., "'Mass defect' tags for biomolecular mass spectrometry," J. Mass Spectrom.; 2003; 38:809-16.

TargetDiscovery
FROM OMICS TO KNOWMICS

Isotope-Differentiated Binding Energy Shift Tags (IDBEST™)

Stable isotope methods like tryptic digestion with [^{18}O]-water,[21] global internal standard technology (GIST™),[22] and isotope-coded affinity tags (ICAT™),[23] are not easily adapted to affinity-MS because the tagged peptides are isobaric with untagged species arising from tryptic digest of the enrichment probe (e.g., antibody or bait protein) or any blocking agents. However, by incorporating a mass defect element into stable isotope paired tags (i.e., IDBEST™), tagged proteins (or the resultant tagged peptides) are shifted by –0.1 amu in the mass spectrum.[24] By using stable isotope paired tags allows a control to be incorporated into every assay, improving the quantitative precision of the assay to the sub-10% level.

IDBEST™ finally allows affinity enrichment of low abundance targets in a low cost protein microarray, standard microplate, or pipette tip chromatography format. In such protocols the IDBEST™ tagged protein and the untagged capture probe (antibody or bait protein) can be digested together and the mass defect tagged peptides quantitatively deconvolved from the untagged chemical noise peptides in the mass spectrum. The resulting mass defect spectrum allows direct comparison of light and heavy tagged peptides in the first MS dimension with any differentially-displayed IDBEST™-tagged peptides trapped and subjected to tandem-MS for identification. Because only one, or a few, target proteins are recovered from a complex sample by the affinity step, there is no need for additional separation of the peptides prior to MS analysis, no matter how complex the bait.

Box B: Stable Isotopes

Normally, an element has a equal number of protons and neutrons. However, when extra neutrons are found in the nucleus of an element, this is referred to as an isotope of the element, and the mass of the isotope is increased by the number of neutrons added. For example the normal version of carbon [^{12}C], atomic mass = 12.000000 g/mol, has 6 protons and 6 neutrons. In nature, however, 1.1% of all carbon atoms contain an extra neutron (i.e., have 6 protons and 7 neutrons), atomic mass = 13.003355 g/mol, and are designated [^{13}C]. Therefore, the formula weight, comprising the weighted sum of all the natural isotopes of carbon, is 12.011037 g/mol.

When too many extra neutrons are added, such as carbon with 6 protons and 8 neutrons ([^{14}C]), the nucleus becomes unstable and the element becomes susceptible to radioactive decay with the release of either alpha, beta, or gamma radiation. For example, [^{14}C] decays (with a half-life of 5568 yr) when one of its neutrons converts to a proton, liberating a high-energy electron (beta particle) to become [^{14}N], atomic mass = 14.006763 g/mol (7 protons and 7 neutrons). The energy of the beta particle is related to the mass defect difference between [^{14}C] to [^{14}N]. Isotopes that do not undergo decay are stable isotopes.

Commonly used stable isotopes include: [^{13}C], [^{15}N], [^{18}O], and [^{2}H] (deuterium). Listings of the most common stable isotopes of all the elements and their natural abundances can be found in various sources.[1-4]

[1] Lee, K. and Anderson, W.A., "Nuclear spins, moments, and magnetic resonance frequencies," in: Handbook of Chemistry and Physics, 60th ed., Weast, R.C., Astle, M.J., eds., pgs. E70-E72 (CRC Press, Boca Raton, FL, 1974).

[2] Heath, R.L., "Table of the isotopes," in: Handbook of Chemistry and Physics, 60th ed., Weast, R.C., Astle, M.J., eds., pgs. B236-B320 (CRC Press, Boca Raton, FL, 1974).

[3] Scientific Instrument Services, Inc , "Exact Masses and Isotopic Abundances http://www.sisweb.com/referenc/source/exactmaa.htm (2004).

[4] JOEL, Inc., Exact Masses of Some Common Elements and Their Isotopes, http://www.jeol.com/ms/docs/elements.html (2004).

---

[21] Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Körner, R., Sönksen, C.P., and Roepstorff, P., "Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using 18O-labeled internal standards," Rapid Comm. Mass Spectrom.; 2000; 14:1226-32.

[22] Chakraborty, A., and Regnier, F.E., "Global internal standard technology for comparative proteomics," J Chrom. A; 2002; 949:173-84.

[23] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R., "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags," Nat. Biotechnol.; 1999; 17:994-9.

[24] Schneider, L.V., Hall, M.P., and Brasseur, M.M., "Mass-defect tagging for proteomic analysis," Genetic Eng. News; 2004; 24:28-30.

TargetDiscovery FROM OMICS TO KNOWMICS

Protein Identification and Top-Down Isoform Characterization

By enabling digestion on the capture surface, and going directly to MS detection, none of the resulting tryptic peptides are lost to sample clean-up (e.g., ZipTip™ or liquid chromatographic) techniques. This provides a level of protein sequence coverage similar to that of top down proteomic methods (Box C), and allows more complete identification of any isoforms that might be present as well as the position of those isoforms along the primary protein sequence.

Previous reports using the mass defect tags in inverted mass ladder sequencing (IMLS™),[29, 25-26] show that the mass defect tag survives MS fragmentation processes, allowing tagged peptide fragments to be discriminated by the mass defect shift from untagged peptide fragments in tandem MS. The N- or C-terminal protein sequence tag derived from IMLS™ yields almost unambiguous protein identication compared to tandem-MS methods (Figure 2). This means that the mass defect tag strategy may be readily extended to true top-down affinity-MS proteomics allowing complete isoform identity and characterization of every target protein without the need for raising monoclonals against every variation.

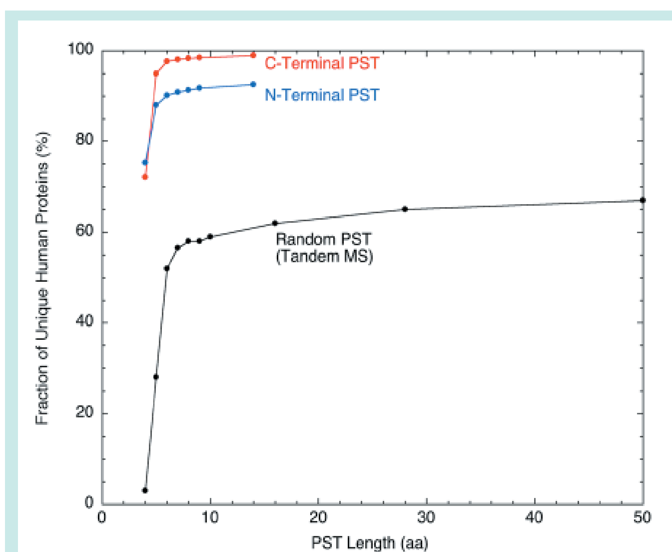

Figure 3. The ambiguity of protein identification as a function of the number of amino acids covered by a protein sequence tag. Bottom-up methods that utilize tandem-MS for sequencing peptides resulting from any position within the protein leave up to 40% ambiguity for human protein assignment.[28] Whereas, sequences derived from the N- or C-terminus of proteins with top-down methods (e.g., IMLS) less than 7% protein assignment ambiguity for human proteins.[26]

Box C: Top-Down & Bottom-up Proteomics

Mass spectrometric proteomic methods are often divided into top-down and bottom-up methods. Most mass spectral proteomics analysis involves bottom-up methods because it is the most compatible with MALDI tandem-MS applications since the single charge state peptides are still small enough to be trapped in the second MS dimension for fragmentation. the disadvantage of bottom-up methods are that many peptides are not recovered, consequently, they are not seen in the mass spectrum, which limits sequence coverage and the ability to discern isoforms.

Top-down methods, pioneered by McClafferty, Kelleher, and coworkers,[1] involve isolation of a multicharge state of an intact protein followed by light fragmention, often by electron capture dissociation, to yield a series of overlapping peptide fragments. When the primary sequence of the protein is known the fragment ions can be reassembled. Fragments that do not match the expected mass from the primary sequence indicate sites of post-translational modifications, mutations, or splice variants (isoforms).

[1] Kelleher, N.L., Top-down proteomics, Anal. Chem.; 2004; 76:197A-203A.

---

[25] Schneider LV, Hall MP, Peterson JN. Methods for sequencing proteins. US6379971 (Apr. 30, 2002).

[26] Hall, M.P., Ashrafi, S., Obegi, I., Petesch, R., Peterson, J.N., and Schneider, L.V., "'Mass defect' tags for biomolecular mass spectrometry," J. Mass Spectrom.; 2003; 38:809-16.

[27] Hernandez, H., Vichchulada, P., Niehauser, S., and Amster, I.J., "Improvements in mass defect labeling for shotgun proteomic analysis," Poster presented at the 52nd Ann. Mtg., Amer. Soc. Mass Spectrom., Nashville, TN (May 23-26, 2004).

[28] Halligan, B.D. and Dratz, E., "The Qcomp method for rapid peptide identification using qualitative amino acid composition and peptide mass," poster and presentation at the Keystone Symposia, Mass Spectrometry in Systems Biology, Santa Fe, NM (Feb. 15-19, 2004).