DRUG DISCOVERY TODAY TARGETS

Stable isotope methods for high-precision proteomics

Luke V. Schneider and Michael P. Hall

Stable isotope tagging methods provide a useful means of determining the relative expression level of individual proteins between samples in a mass spectrometer with high precision (coefficients of variation less than 10%). Because two or more samples tagged with different numbers of stable isotopes can be mixed before any processing steps, sample-to-sample recovery differences are eliminated. Mass spectrometry also allows post-translational modifications, splice variations and mutations (often unnoticed in immunoassays) to be detected and identified, increasing the clinical relevance of the assay and avoiding the issues of non-specific binding and cross-reactivity observed in immunoassays. Several stable isotope tagging methods are available for use in proteomics research. We discuss the advantages and disadvantages of each technique with respect to biomarker discovery, target validation, efficacy and toxicology screening and clinical diagnostic applications.

Now that we know that gene expression does not always correlate with protein expression [1], proteomics – the large-scale, high-throughput identification and quantification of proteins – is playing an increasingly important role in drug discovery and development. Proteomics is used for the discovery and validation of new therapeutic or diagnostic targets (biomarker discovery), the efficacy and toxicology screening of new drug candidates, and patient diagnostics or profiling before therapy selection (theranostics). Recent developments in stable isotope protein tagging technology and have made mass spectrometric detection applicable to both global profiling (an open method) and affinity enrichment (a closed method) proteomics. Open methods, such as global profiling, have proven useful for biomarker discovery, where the research objective is to find proteins associated with a disease state [2] or toxicological reaction [3]. Closed methods, such as immunoassays (IAs) or protein–protein interaction (baiting) strategies, enrich low-abundance proteins so that they can be detected; they are particularly useful in efficacy or toxicology screenings and patient profiling (theranostics), and form the backbone of the $11.2 billion *in vitro* diagnostics industry [4].

A major issue in global proteomics methods is the limited dynamic range of the detection methods (Figure 1). Proteins exhibit a very wide range in concentrations, with a dynamic range of 10^5 in bacteria [5] to 10^7–10^8 in human cells [6] to at least 10^{12} in plasma [7]. Since there is no technique to amplify low-abundance proteins comparable to the polymerase chain reaction for nucleic acids, both gel and mass spectrometry (MS) methods often fail to detect low-abundance proteins [8]. For example, proteomic characterization of nipple aspirate fluid (NAF) of the healthy breast using MudPIT [9] failed to reveal kallikrein 3 (PSA). Yet IAs show that PSA is present at significant concentrations in NAF from the healthy breast [10]. Baiting (an affinity-enrichment-MS strategy) can be used to identify biomarker

Luke V. Schneider*
Michael P. Hall
Target Discovery, Inc.,
4015 Fabian Way,
Palo Alto,
CA 94303,
USA
*e-mail: luke_schneider@targetdiscovery.com

1359-6446/04/$ – see front matter PII: S1359-6446(04)03381-7

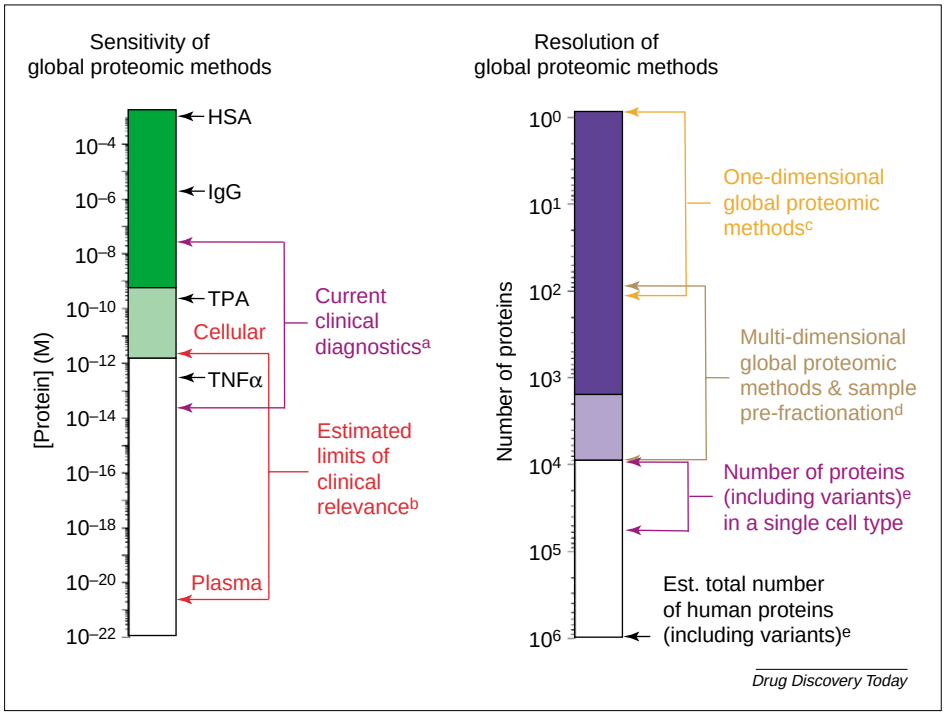

FIGURE 1

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 [46]. [b]The limits of clinical relevance are calculated based on one protein copy in the volume of a mammalian cell (≈1 x 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 variants includes splice variants, mutations and post-translational modifications. The colored bars denote the current detection or resolution limits, adapted from Kenyon *et al.* [8]. The plasma concentrations for human serum albumin (HSA), immunoglobulin G (IgG), tissue plasmogen activator (TPA), and tumor necrosis factor alpha (TNFα) are shown for reference.

function and to associate the biomarker with a disease pathway through identification of other proteins with which it interacts, a targeted biomarker discovery activity.

Mass spectrometric approaches are often further divided into top-down and bottom-up methods (Figure 2, Box 1). Top-down is a term used to describe the analysis of the intact protein by MS with the goal of characterizing mutations, splice variations and post-translational modifications (protein variants). Bottom-up methods involve the digestion of the intact protein before mass spectrometric analysis of the resulting peptides. The primary goal here is to determine the relative expression of the parent protein between two samples (differential display).

Affinity-enrichment-MS approaches have been successfully applied to top-down proteomics [11–15], but have not been quantitative unless coupled to an additional assay (e.g. surface plasmon resonance) [16–17]. Recent advances in stable isotope ratio tagging technology appear to have finally enabled quantitative (stable isotope) affinity-enrichment-MS assays in both top-down and bottom-up proteomic approaches.

Stable isotope ratio mass spectrometry for proteomics

Quantifying protein levels across samples is often difficult in any proteomic method that does not utilize internal standards because of protein recovery differences during sample preparation. This is particularly true in MS because of irreproducible fluctuations during ionization and ion competition effects. However, a unique feature of mass spectrometry is its ability to quantitatively discern the relative abundance of stable isotopes (Box 2) in otherwise identical chemical species, irrespective of competing ion concentrations. This feature is exploited in a series of protein tagging technologies, which are rapidly making mass spectrometry the detection method of choice for quantitative proteomics.

While absolute abundance measurements can vary widely from sample to sample and day to day in mass spectrometry, variability in relative abundance (i.e. where the counts are scaled relative to the highest-abundance peak in the mass spectrum) has been observed to be <20% [18] and as low as 10% [2] for the Ciphergen ProteinChip®. Stable isotope methods circumvent the problem of varying sample recovery. A stable isotope version of a chemical exhibits the same ionization efficiency as its lighter counterpart because they are essentially the same chemical entity. Except for deuterium [19–20], most stable isotopes do not impart recovery differences through other analytical methods (e.g. chromatography or electrophoresis), but are easily distinguished and quantified in mass spectra.

Stable isotope ratio mass spectrometry (SIRMS) has a long history in nutritional science [21–22]. Hellerstein first applied the *in vivo* incorporation of stable isotopes to measure the rates of protein synthesis and destruction [23], a technique ultimately adapted by Ong *et al.* for the stable isotope labeling by amino acids in cell culture (SILAC) for protein differential display. [24] Gygi *et al.* suggested the use of stable isotope chemical tags (*ex vivo*) as a way to perform protein differential display, coupling this with affinity enrichment of the SIRMS tag itself [25]. SIRMS approaches using chemical tagging are the focus of this review because of their widespread applicability to any sample. SIRMS tagging for proteomics is rapidly evolving and we review each of the various SIRMS proteomic technologies in the following sections, discussing their advantages and limitations for various proteomic applications (e.g. biomarker discovery, target validation, efficacy and toxicology screening, patient profiling and clinical diagnostics).

SIRMS technologies for global proteomics

[^{18}O]-Water

The pioneering SIRMS proteomic method, the use of a 1:1 mixture of [^{16}O]:[^{18}O] water during protein digestion, was originally used to determine the C-terminal peptide after digestion [26–27]. It was adapted for protein differential display by performing a tryptic digestion of one sample with [^{16}O] water and a second sample with [^{18}O] water [28]. The two digests were mixed and the relative abundances of individual peptide pairs determined by MS (Figure 3). A disadvantage of this approach is the slow back exchange of [^{16}O] water and [^{18}O] water with the terminal isotope-labeled hydroxyl group s once the two digests are mixed. However, both trypsin-catalyzed back exchange and pH-mediated exchange can be suppressed for up to 24 h with the addition of 1–5% formic acid [29]. To avoid problems with back exchange, samples must be processed quickly to obtain the best quantitative results. Since the C terminus is only labeled by back exchange with [^{18}O] water, the C-terminal peptide typically remains undetected when the [^{18}O]-water methodology is applied in global profiling applications.

Global internal standard technology

In an effort to design a global proteomic tagging system, Regnier and co-workers [30–31] developed global internal standard technology (GIST) (Figure 3). GIST involves the proteolytic digestion of control and experimental samples, tagging of the resulting peptides (usually with deuterated and nondeuterated versions of acylating agents) and mixing of the labeled peptides, followed by quantification by mass spectrometry [30]. The strategy is frequently combined with liquid chromatography (e.g. RP-HPLC) to reduce the complexity of the resulting mass spectrum. Almost all peptides in a complex mixture have been shown to be reproducibly labeled with deuterated ($[^2H]_3$) and non-deuterated ($[^1H]_3$) versions of *N*-acetoxysuccinimide (NAS), which labels primary amino groups (i.e. N termini and lysine sidechains) [30]. Only the N-terminal peptides of proteins blocked at the N-terminus not containing lysine will not be labeled. For this reason, Regnier suggests trypsinization in the presence of [^{18}O] water for either the control or the experimental sample before acylation of the primary amino groups, since the C termini of N-terminal-blocked peptides will still be differentially labeled and the N terminus of the C-terminal peptide (which is not [^{18}O]-labeled) will still be acylated [32]. With this additional step, the technology theoretically encodes all the peptides from any complex protein sample.

An issue with the GIST is the variable difference between the masses of light and heavy tagged peptides. The mass difference in the paired peptides (i.e. heavy versus light) depends on the number of acylated sites in the peptide, which varies from 3 to 13 amu for single charge states [32], making automated peak-finding problematic in the absence of sequence information. On the other hand, the specific mass shift, once identified, can give vital information regarding the number of lysines and position of the peptide within the protein (i.e. N-terminal, internal or C-terminal), which can lead to faster and less ambiguous protein identification.

Another issue is the use of [^{1}H]/[^{2}H] isotopes and their effect in RP-HPLC. It is apparent that significant substitution of [^{1}H] for [^{2}H] in most tags changes the retention times of labeled species in RP-HPLC, which confounds the MS analysis, resulting in inferior quantification [19]. The use of [^{12}C]/[^{13}C] instead of [^{1}H]/[^{2}H] in the coding tags eliminates this problem entirely [33–34].

Because GIST acylation tags also react with primary amino groups (e.g. lysine residues), they can reduce ionization efficiency in positive ion mode for many peptides [19]. This problem can be circumvented by the incorporation of a non-nucleophilic basic site or other site of positive charge into the acylation tags, such as [3-(2,5-dioxopyrrolidin-1-yloxycarbonyl)-propyl]-trimethylammonium chloride or reactive imidazoles ($[^1H]_9$/$[^2H]_9$ versions) [19,35,36].

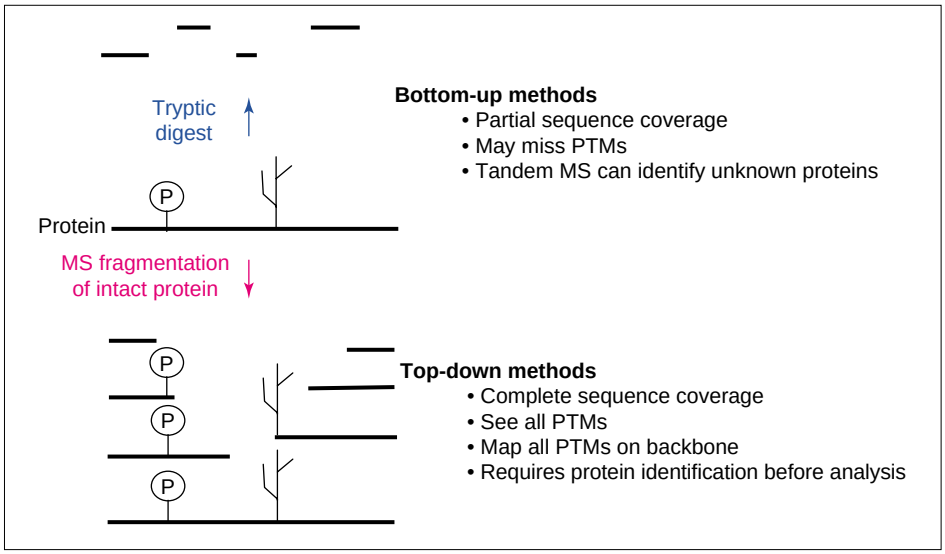

FIGURE 2

Schematic drawing (adapted from Kelleher, [56]) illustrating the difference in amino acid sequence coverage between top-down and bottom-up mass spectrometry (MS) proteomic methods. In bottom-up methods (top panel) the protein is digested and the resulting peptides chromatographed (for desalting and sample simplification), involving a possible loss of peptides prior to MS analysis. The protein is identified by tandem MS fragmentation. In top-down methods the intact protein is ionized and fragmented in the mass spectrometer such that all fragments are represented in the resulting spectrum. Therefore, top-down methods are more likely to see post-translational modifications (PTMs), such as phosphorylations or glycosylations (represented schematically in the diagram).

Isotope tags for relative and absolute quantification (iTRAQ™)

A common complaint about global stable isotope profiling strategies is the time

BOX 1

Top-down versus bottom-up proteomics

Mass spectrometric proteomics are often divided into top-down and bottom-up methods (Figure 2). Bottom-up methods involve the digestion of the intact protein before mass spectrometric analysis of the resulting peptides. Most of the protein analysis by mass spectrometry involves bottom-up proteomics. The advantage of bottom-up proteomics is that the masses of single charge state species are small enough to be used for fragmentation and identification of the parent protein (tandem MS sequencing) [54]. Another advantage is that more accurate masses can be assigned to the peptides because their smaller mass-to-charge ratio places them in a higher-resolution region of the mass spectrometer. This facilitates peptide mass mapping for protein identification [55]. A disadvantage of bottom-up methods is that some peptides are not recovered during sample clean up, are too small or poorly ionized and are consequently not seen in the mass spectrum. This limits sequence coverage and the ability to discern protein variations. Furthermore, in global proteomic profiling, the separation challenge is increased because the average 45 kDa protein yields 35 tryptic peptides.

Top-down is a term used to describe the analysis of the intact protein by mass spectrometry [56]. Top-down methods, [56–57] involve isolation of one or more charge states of an intact protein. The intact protein is lightly fragmented to yield a series of overlapping peptide fragments. When the primary sequence of the protein is known, the fragment ions can be reassembled, much like the reassembly of a restriction digest map of DNA. Fragments that do not match the mass expected from the primary sequence indicate sites of post-translational modifications, mutations or splice variants. Because of the overlaps in the peptide fragments, the specific sites of such modifications can be pinpointed. The disadvantages of the top-down method are, first, that the protein primary sequence must be known to reassemble the fragment ion map; second, that the method typically requires an expensive Fourier transform inductively coupled resonance (FTICR) mass spectrometer; and third, a single chemical species is spread over multiple charge states, limiting sensitivity and increasing the sampling time required.

Recently, a mass defect tag method for determining the N- and C-terminal amino acid sequence of intact proteins, called inverted mass ladder sequencing (IMLS™) [51], has been reported, which may lead to easier application of top-down methods to unknown proteins.

BOX 2

Stable isotopes

Normally, an element has an 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 six protons and six neutrons. In nature, however, 1.1% of all carbon atoms contain an extra neutron (i.e. have six protons and seven 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 six protons and eight neutrons ([^{14}C]), the nucleus becomes unstable and the element becomes susceptible to radioactive decay with the release of α, β or γ 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 (β particle) to become [^{14}N], atomic mass = 14.006763 g/mol (seven protons and seven neutrons). Isotopes that do not undergo decay are called stable isotopes. Commonly used stable isotopes include [^{13}C], [^{15}N], [^{18}O] and [^{2}H] (deuterium).

required for separation of the peptides to prevent confounding overlaps in the mass spectrum. ABI (www.appliedbiosystems.com) recently announced iTRAQ™, which is potentially more cost-effective than GIST because up to four samples can be screened simultaneously with four isotopically distinguishable reagents (Figure 3) [37]. The iTRAQ methodology utilizes isobaric tags containing both reporter and balancer groups. The reporter is quantitatively cleaved during collision-induced dissociation (CID) to yield an isotope series representing the quantity of a single peptide of known mass from each of up to four different samples. This quantification group (the reporter) is 'balanced' by a second group (the balancer) depleted of the same stable isotopes, which maintains each isotopic tag at exactly the same mass. Since the peptide remains attached to the isobaric tags until CID is conducted, the peptide is simultaneously fragmented for sequence identification.

The current generation of iTRAQ reagents labels lysine residues and the N termini of peptides, meaning that most peptides are multiply labeled (as with GIST). Therefore, iTRAQ suffers the same peptide overabundance problem and must be coupled with one or more dimensions of chromatographic or electrophoretic separation before MS analysis to limit the number of isobaric tagged peptides in the first MS dimension. The advantage of iTRAQ over these methods is that the label is cleaved in the tandem MS before quantification. This means that competing untagged isobaric peptides do not interfere with quantification as they do in GIST or [^{18}O]-water methods.

Because differences in peptide levels can only be determined after tandem MS, the first MS dimension cannot be used to pre-screen peptides for differential expression before tandem MS identity determination. Therefore, each and every peptide must be subjected to tandem MS analysis, making iTRAQ both time-consuming and sample-intensive for biomarker discovery applications. Furthermore, any untagged isobaric chemical noise may confound tandem-MS sequencing of the iTRAQ labeled peptides.

Another issue with this method is the problem of protein variants. Any variant of the peptide of interest will not be isobaric with the same tagged peptides from control samples. Such non-isobaric peptides can be detected by their absence, but may be falsely interpreted as down-regulation of the parent protein. Furthermore, such peptides may be isobaric with other peptides, confounding

the interpretation of expression levels or sequences of other peptides. However, in target validation, patient profiling or toxicological screening applications, where the masses of the peptides are known, iTRAQ™ is potentially very cost-effective.

As bottom-up proteomic methods, GIST, iTRAQ and [^{18}O]-labeling strategies are designed to isotopically encode virtually all of the peptides from a protein digest. However, with all these strategies, the very large number of labeled peptides resulting from clinical tissue samples (up to 350,000) very easily exceeds the resolution capacity of the MS, even when coupled to multi-dimensional liquid chromatographic separations. Thus, peptide quantification is confounded by the co-elution of isobaric peptides [38]. Most examples in the literature utilize GIST in combination with affinity techniques to substantially reduce the complexity of the peptide mixtures and allow more accurate differential display quantification, such as the enrichment of histidine-containing GIST peptides using an immobilized metal affinity column (IMAC) and cysteine-containing GIST peptides by reversible covalent attachment to thiopropyl Sepharose 6B resin [38–39]. These peptide-selection techniques are equally applicable to iTRAQ and [^{18}O]-water global proteomic strategies. It is apparent that, for meaningful data to be generated, affinity enrichment followed by multidimensional chromatography to reduce sample complexity is essential for any global peptide labeling strategy used to analyze a complex proteome.

Isotope-coded affinity tags (ICAT™)

An alternative method to reduce the complexity of protein mixtures is the ICAT™ strategy developed by Aebersold and co-workers [25,40]. ICAT selectively targets cysteines, but can be expanded to any residue to which a tag can be conjugated (Figure 3). The tag contains an iminobiotin moiety for affinity clean-up prior to subsequent HPLC separation(s) and MS analysis. Therefore, only peptides that contain cysteine are analyzed.

With ICAT, protein samples from healthy and diseased (or perturbed) sources are denatured, reduced and labeled

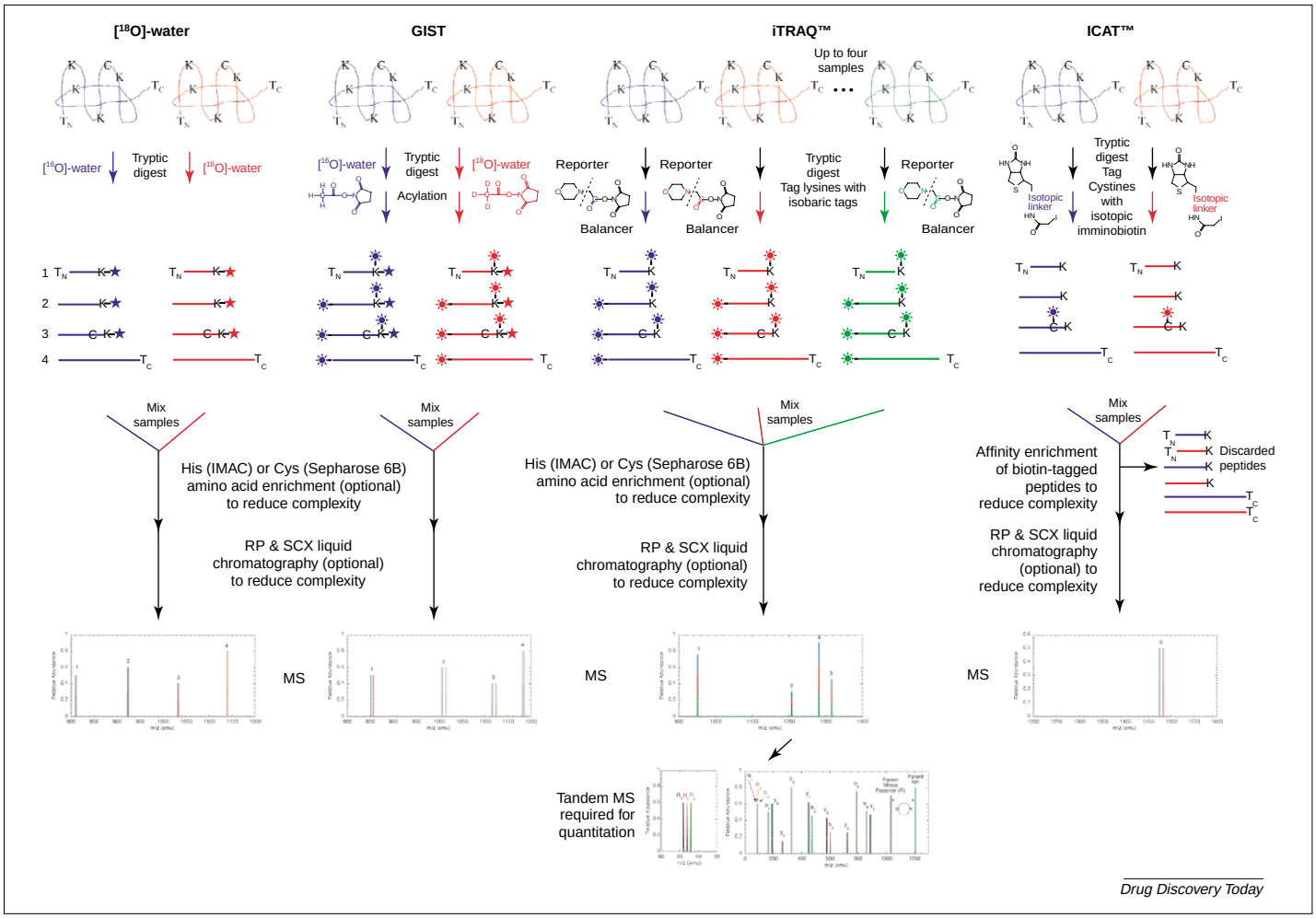

FIGURE 3

The global profiling techniques: [^{18}O]-water, GIST, iTRAQ and ICAT. Stable isotope tagging techniques used for global proteomic profiling include: [^{18}O]-water labeling through tryptic digestion, global internal standard technology (GIST), isotope tags for relative and absolute quantification (iTRAQ), and isotope-coded affinity tags (ICAT). The stars indicate the position of the label on the peptides. The lighter-isotope-encoded sample is shown in blue and the heavy-isotope-encoded sample is shown in red. Representative mass spectra (MS) of the mixed labeled peptide samples are correspondingly color coded. T_N and T_C indicate the protein N- and C termini, respectively. K and C indicate lysine and cysteine residues, respectively.

separately. All cysteines in the healthy sample are modified with one isotopic version of the tag, and the cysteines in the perturbed sample are tagged with the opposite isotopic reagent. The protein samples are then mixed and proteolyzed. Cysteine-tagged peptides are enriched by affinity chromatography of the iminobiotin tag through an avidin column and are subsequently chromatographed by RP-HPLC, alone or in combination with ion exchange liquid chromatography such as strong cation exchange (SCX), and introduced into a mass spectrometer for quantification of expression differences. Further analysis of differentially expressed peptides by tandem MS allows sequencing and identification of any differentially expressed peptides.

ICAT has been successfully used in several applications; these include the determination of microsomal protein levels in human myeloid leukemia cells undergoing pharmacologically induced differentiation [41] and the identification of changes in the protein composition of lipid rafts isolated from control and stimulated Jurkat T cells [42], and cleavable ICAT (cICAT) reagents have been used to identify protein expression differences in cystic fibrosis [43].

The original [^{1}H]- and [^{2}H]-isotopic ICAT tags led to variable retention times in RP-HPLC for the same peptide [33]. This issue has been rectified in the second-generation ICAT reagents, which incorporate [^{12}C] and [^{13}C] isotope pairing. The iminobiotin affinity tag can also be problematic. Excess label or any endogenous biotin in the sample matrix (e.g. serum) may reduce the affinity column capacity because of competition for binding sites. Furthermore, ICAT peptides are sometimes isobaric with non-tagged peptides eluting from the affinity chromatography step, leading to false-positive and false-negative detections.

The ICAT tag itself can lead to collisional energy losses during fragmentation leading to sequencing difficulties [44]. The third-generation (cICAT) ICAT reagents contain an acid-cleavable group between the biotin and the isotopically labeled linker so that a smaller attachment group is left on the cysteine residue after acidification, allowing more robust fragmentation and sequencing analysis [45].

A serious disadvantage of any peptide-selective strategy (e.g. ICAT or cysteine- or histidine-selective enrichment in either GIST or [^{18}O]-water methods) is the potential under-representation of proteins. In a recent review, Zhang *et al.* estimated 96.1% protein coverage for Cys tags and 97.3% theoretical coverage for His tags in human proteins [45]. Therefore, most proteins in any mixture would be represented by at least one cysteine- or histidine-containing peptide. However, even for proteins that contain cysteine or histidine, there is the strong possibility that tryptic fragments containing these residues may not ionize efficiently in the mass spectrometer or may not be recovered through the affinity or HPLC chromatography, thus leading to populations of proteins that are not represented in ICAT analysis, or in peptide-selective GIST or [^{18}O]-water analyses. Even with the low relative abundances of cysteine and histidine, the number of peptides from a tissue sample must still be further reduced by one or two dimensions of HPLC before MS analysis.

An even greater problem resulting from peptide-selective ICAT, GIST or [^{18}O]-water strategies is a loss of peptide redundancy for a given protein species. Any protein variation affecting the enriched peptide, or its recovery during sample preparation, may result in a false-positive or false-negative detection event since the expected peak pair would not be seen. Since only one or a few peptides (those containing the relatively low-abundance cysteine or histidine) are used as representatives of the protein, there may also be important, clinically relevant structural features, such as post-translational modifications, that are completely missed unless these features are located in the cysteine- or histidine-containing peptide subset.

SIRMS technologies for affinity-enrichment methods

A critical limitation of all global proteomic profiling methods is the dynamic range of concentrations of proteins (and resultant peptides) in tissue samples, which has been argued to be as much as 12 [7] to 17 [8] orders of magnitude (i.e. 1 protein copy per ml of plasma compared to 10^{17} copies of serum albumin per ml [46]) in human plasma. Mass spectrometers have a very limited dynamic range (two to three orders of magnitude) by comparison. While subcellular fractionation, high-abundance protein depletion and peptide separations applied before MS analysis may extend the effective dynamic range by a few orders of magnitude [8], global proteomic studies often fail to identify known biomarkers because they are present at too low an abundance to be detected by MS. The solution to this dilemma is a targeted proteomic strategy that utilizes affinity or baiting strategies to selectively enrich lower-abundance proteins. Such affinity enrichment before MS analysis has only successfully been reported for intact proteins [12], with quantification of the target protein requiring the use of some other spectrometric method [16].

Bottom-up global profiling strategies also require reassembly of the target protein from those peptides seen in the MS, which leaves significant gaps in the sequence coverage and misses many protein variations, particularly where cysteine- or histidine-residues are selectively enriched. Such reassembly can yield ambiguities in protein identification even with extensive sequence coverage by tandem MS [47]. By targeting the intact protein instead of a peptide, the entire amino acid sequence is covered in the MS, even if the protein is digested on the enrichment surface before MS analysis. Since global profiling strategies typically require some amount of chromatographic separation to reduce the sample complexity, they are more easily coupled to electrospray MS. Affinity capture of tagged proteins, particularly in a microarray format, is more easily coupled directly to MALDI-MS because the sample complexity is significantly reduced, even with on-chip digestion.

BOX 3

The mass defect

Although 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 as a result of the efficiency with which these nucleons are packed together. The amount of energy liberated depends on the number of nucleons packed together into the nucleus (i.e. 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 the mass expected based on the number of nucleons that comprise its nucleus (Figure i). By convention, this mass defect is defined as zero for ^{12}C and the mass defects of all other elements are scaled to this standard. This mass difference is what accounts for the energy released upon nuclear fission or fusion. For example, combining two deuteriums (4.028204 Da) to form [^{4}He] (4.002603 Da) liberates 0.025601 amu as energy (1.10^{12} J/mol [^{2}H]).

Elements commonly found in biomolecules (e.g. C, H, N and O) have negligible mass defects. However, the mass defect of those 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 [48]. With sufficient mass spectrometric resolution, mass defect tagged peptides can be quantitatively deconvolved from non-tagged peptides. This principle underpins IMLS, a top-down sequencing method [51,58]. When stable isotopes are used in conjunction with a mass defect element in the tag, this forms the basis of the isotope-differentiated binding energy shift tags (IDBEST) method [49–50]. 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) [52] and metal-coded affinity tags (MeCAT) [59].

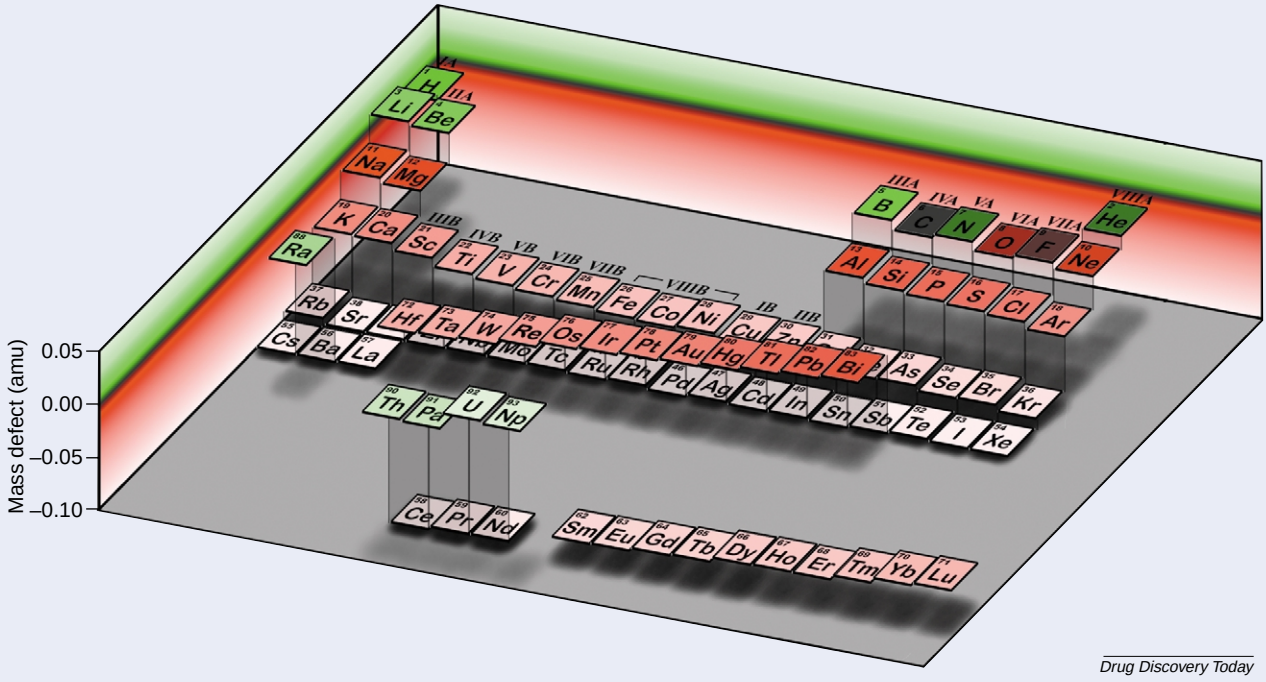

FIGURE i

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 [50].

Isotope-differentiated binding energy shift tags (IDBEST™)

Stable isotope methods like [^{18}O] water, GIST and ICAT are not easily adapted to affinity-enrichment-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 blocking agents. However, two recent advances in stable isotope tagging technologies may finally enable quantitative affinity-MS methods in either bottom-up or top-down strategies. By incorporating a mass defect element (Box 3) 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 from all untagged peptide species [48–50]. The ability to distinguish mass-defect-tagged species from untagged chemical noise has been independently demonstrated by several groups [51–53], including the quantitative recovery of the tryptic peptides of stable-isotope-paired mass-defect-tagged bovine serum albumin from *Escherichia coli* whole cell lysate tryptic digest with a 4% quantitative error, [49] and recovery of IDBEST-labeled PSA (labeled on primary amino groups) from human serum (Figure 4).

IDBEST allows affinity enrichment of low-abundance targets in a low-cost protein microarray or a standard microplate or in pipette tip chromatography assay formats.

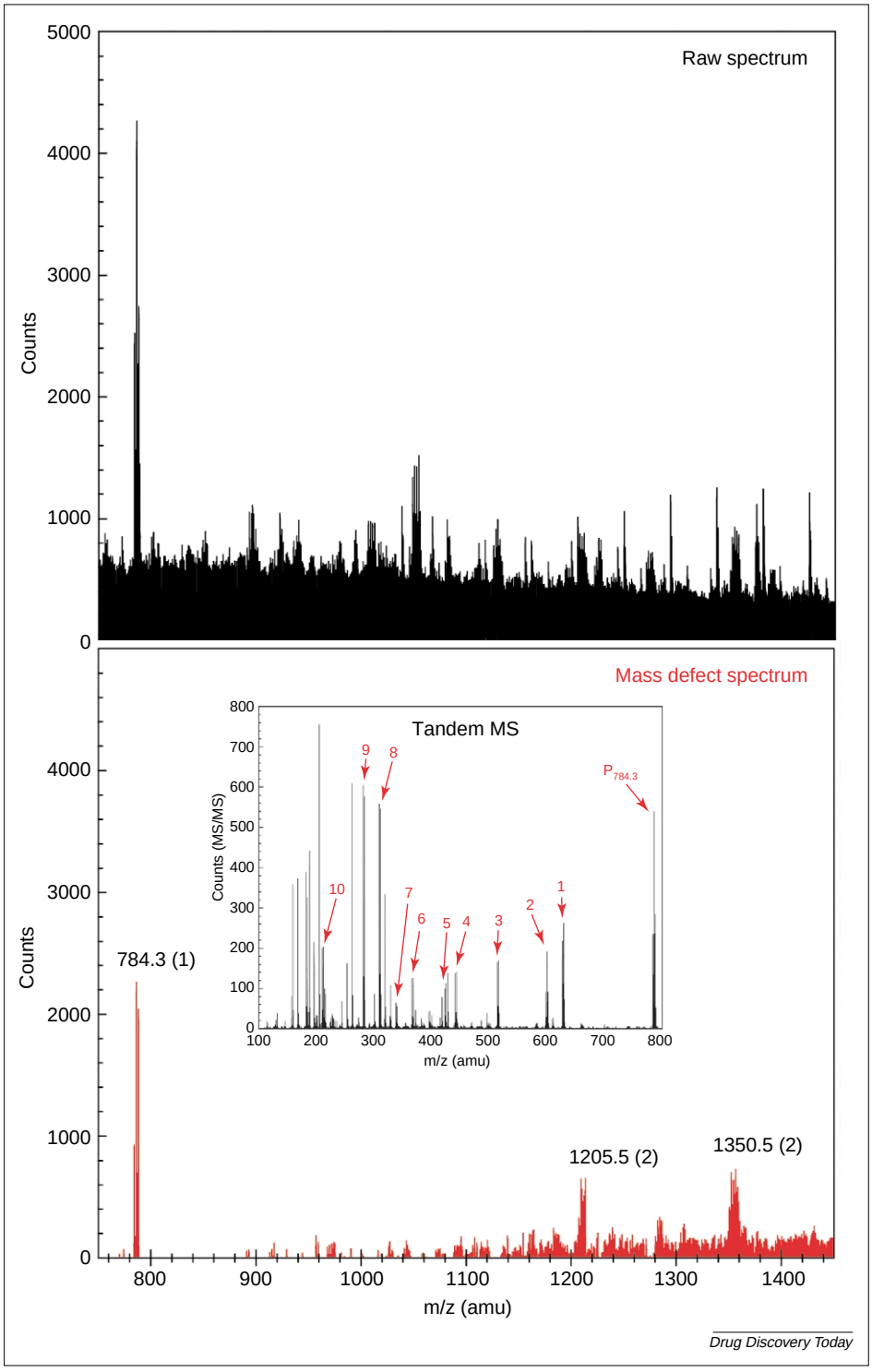

FIGURE 4

Recovery of prostate specific antigen (PSA) present at 0.1 mg/mL in human serum by affinity enrichment (monoclonal antibody bound to agarose beads). The plasma sample was labeled with the lysine reactive IDBEST reagent, 3-bromo-1-(5-carboxy-pentyl)-pyridinium bromide-NHS ester (BDPOPP) at a 20:1 molar ratio to lysine. The labeled serum was incubated with mouse anti-PSA mIgG bound to agarose beads. The complex was washed and digested directly on the beads with both trypsin and chemotrypsin. The supernatant was recovered and desalted by reverse phase chromatography (C18 column), lyophilized and resuspended with α-cyano-4-hydroxycinnamic acid (CHCA), spotted and analyzed on an MDS Sciex QSTAR®. The raw data was deconvolved with the IDBEST software to eliminate the spurious IgG peaks and identify the mass defect peaks indicated (number of labels). The isotope series (5 amu) corresponding to the 784.3 amu monoisotopic peptide was collected and subjected to collision-induced dissociation (CID). The corresponding tandem MS sequence is shown in the inset, and identified as the N-terminal PSA peptide (IVGGW). The numbered peaks in the inset correspond to the parent labeled peptide ($P_{784.3}$), the parent peptide minus the Br-pyridine (Br-Pyr) moiety of the label (1), the labeled b_4-ion (2), the y_4-ion minus the Br-Pyr moiety (3), the b_4-ion minus the Br-Pyr moiety (4), the b_4-ion minus the Br-Pyr and H_2O (5), the b_3-ion minus the Br-Pyr (6), the b2-ion minus the Br-Pyr (8), and the b1-ion minus the Br-Pyr (10). Peaks 7 and 9 correspond to isotopically-labeled internal fragments VGG and VG minus the Br-Pyr moiety.

In such formats 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 (Figure 5). 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 (inset, Figure 4). 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 before MS analysis, no matter how complex the bait. Previous reports using the mass defect tags in inverted mass ladder sequencing (IMLS™) [51] show that certain mass defect tags survive MS fragmentation processes, allowing tagged peptide fragments to be discriminated by the mass defect shift from untagged peptide fragments in tandem MS, thus eliminating the problem of capturing nearly isobaric chemical noise in the collision cell during tandem MS.

A potential problem with the mass defect tag approach is the natural abundance of S and P in peptides. Both S and P exhibit a –0.05 amu mass defect; therefore, multiple S or P atoms in a peptide will cause that peptide to be isobaric with mass-defect-tagged peptides. Br-containing mass defect tags overcome this issue by creating a matching isotope series at both the [^{79}Br] and [^{81}Br] positions, providing a means of readily identifying false-positive mass defects. [51] Furthermore, a double mass-defect tag has been reported [53] which increases the ability to resolve the tagged peptides by –0.2 amu and creates a unique Br-triplet series for each tagged species. The probability of any peptide containing more than two S or P atoms is low.

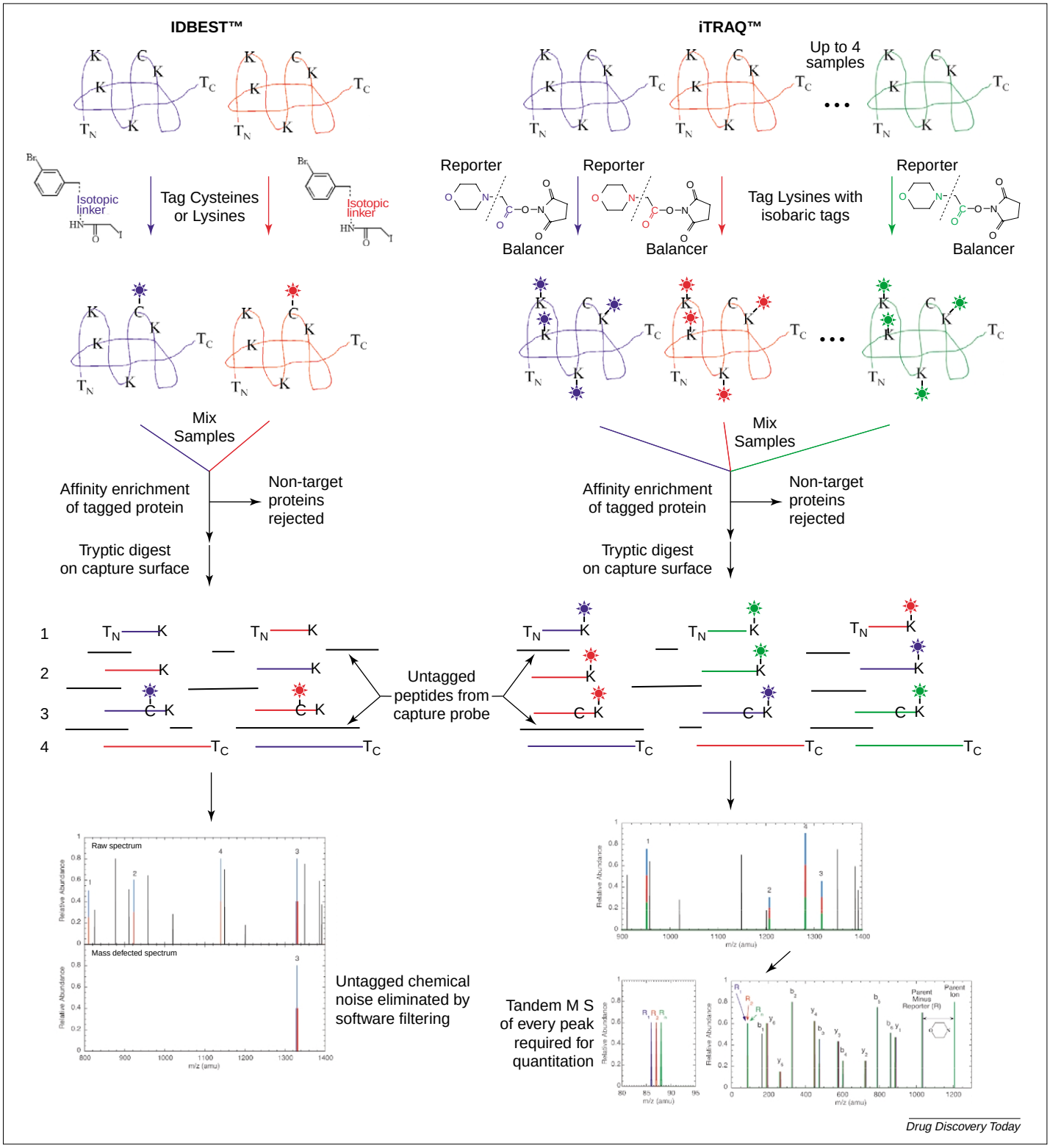

FIGURE 5

Affinity enrichment SIRMS technologies: IDBEST and iTRAQ. Stable isotope tagging techniques used for affinity enrichment proteomic methods include: isotope tags for relative and absolute quantification (iTRAQ) and isotope-differentiated binding energy shift tags (IDBEST). The stars indicate the position of the label on the peptides. The lighter-isotope-encoded sample is shown in blue and the heavy-isotope-encoded sample is shown in red. Representative mass spectra (MS) of the mixed labeled peptide samples are correspondingly color coded. T_N and T_C indicate the protein N- and C-termini, respectively. K and C indicate lysine and cysteine residues, respectively.

iTRAQ for affinity-MS

iTRAQ reagents also have the potential to be used in similar affinity-enrichment-MS strategies. Since the iTRAQ-tagged peptides are only detected and quantified by tandem MS, it should be possible to screen every peptide peak seen in the first MS dimension by tandem-MS to identify those

peptides resulting from the iTRAQ-tagged target protein(s). The potential disadvantages of iTRAQ in this approach are, first, that every peptide must be screened by tandem MS, increasing sample analysis time and the amount of sample needed, and, second, that isobaric peptides resulting from the capture probe can confound sequencing of the tagged peptides. A monoclonal antibody will produce 120 tryptic peptides on average. Even if only half are seen in the MS, and we assume that only 1 min per peptide is required for tandem MS analysis of each peptide, this means at minimum of 1 h of continuous analysis per sample is needed to identify and quantify all the iTRAQ-labeled peptides from a single MALDI spot. This is not practical since the sample will be completely consumed within a few minutes. Furthermore, if an iTRAQ-tagged peptide is isobaric with another peptide, both will be trapped during tandem MS sequencing. The only difference between the resulting tandem MS fragments arising from each peptide will be the small balancer residue left behind after cleavage of the iTRAQ reporter, making it difficult to distinguish which fragment ions result from which peptide. These issues may restrict the usefulness of iTRAQ for targeted biomarker discovery applications, but are balanced by the increased throughput made possible by the four-sample multiplexing.

Either IDBEST or iTRAQ would be suitable for low-cost target validation assays and toxicological or efficacy screening, where the target mass(es) are known ahead of time. However, iTRAQ would require a tandem MS to perform the analysis, whereas IDBEST would only require a single stage MS (about half the instrumentation cost). IDBEST would appear to be the only practical solution for targeted biomarker discovery applications since the time and sample amounts needed for tandem MS analysis of every iTRAQ peptide would be prohibitive.

Conclusions

Although a relatively recent addition to the proteomics scene, SIRMS proteomic methods have proven themselves as highly quantitative tools with the ability to unambiguously identify both the protein and its variants and to overcome the sample-to-sample recovery variabilities associated with non-SIRMS MS-proteomic methods (e.g. MudPIT and the Ciphergen ProteinChip®). Stable isotope proteomic methods have until recently been limited to global profiling strategies primarily focused on biomarker discovery applications. One major challenge of global profiling methods has been the need for sample simplification by affinity chromatography of the tag (e.g. ICAT and cICAT) or selective enrichment of peptides containing low-abundance amino acids, and/or followed by extensive liquid chromatography of the tagged peptides (ICAT, GIST, [^{18}O]-water, and iTRAQ). These pre-MS complexity-reducing methods are limited to biomarker discovery applications as they are too expensive and time-consuming to be used for target validation, toxicology and efficacy screening, or patient diagnostics. The other major challenge of global profiling strategies has been the ability to reach down into the proteome for low abundance biomarkers, which has limited the utility of global profiling even for biomarker discovery.

Affinity-enrichment-MS methods are the logical extension of stable isotope ratio proteomics, but have not been practical until recently because of the inability to sort bait from target peptides in the mass spectrum. Affinity-enrichment-MS now appears be in reach with the commercial release of both IDBEST and iTRAQ. While iTRAQ has not yet been applied to affinity-enrichment-MS applications, Figure 4 illustrates the utility of IDBEST in these applications. By enriching low-abundance proteins or protein classes, the number of proteins (and resulting peptides) is reduced to a number manageable within the resolution of the mass spectrum without the need for additional pre-separation. High-throughput affinity-enrichment-MS in microarray, pipette tip and microtiter plate formats is time- and cost-effective for target validation, toxicology and efficacy screening, and patient profiling applications. In addition, either baiting or affinity enrichment strategies can be used for targeted biomarker discovery applications to drill down to lower-abundance proteins that have so far eluded global profiling strategies.

Acknowledgements

Figure 4 presents previously unpublished data of the authors and co-workers William P. Chang, John P. Wilson, Robert Petesch, Lane A. Clizbe, and Siamak Ashrafi (Target Discovery, Inc.). The MS and tandem-MS data was generated by Protana Analytical Services (Toronto, Canada). This work was in part supported by the National Cancer Institute under grant # G1R43 CA 103085–01.

References

1 Gygi, S.P. *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730
2 Petricoin, E.F. *et al.* (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577
3 Steiner, S. and Anderson, N.L. (2000) Expression profiling in toxicology – potentials and limitations. *Toxicol. Lett.* 112-113, 467–471
4 Frost and Sullivan (2002) World *in vitro* Diagnostic Technologies Market. A116-53, Frost & Sullivan
5 Ingraham, J.L. *et al.* (1983) Growth of the Bacterial Cell. Sinauer Associates
6 Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19, 1853–1861
7 Corthais, G.L. *et al.* (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 21, 1104–1115
8 Kenyon, G.L. *et al.* (2002) Defining the mandate of proteomics in the post-genomics era. *Mol. Cell. Proteomics* 1, 763–780
9 Varnum, S.M. *et al.* (2003) Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. *Breast Cancer Res. Treat.* 80, 87–97
10 Sauter, E.R. *et al.* (1996) Prostate specific antigen levels in nipple aspirate fluid correlate with breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 5, 967–970
11 Nedelkov, D. *et al.* (2004) High-throughput comprehensive analysis of human plasma proteins: a step toward population proteomics. *Anal. Chem.* 76, 1733–1737

12 Nelson, R.W. *et al.* (1995) Mass spectrometric immunoassay. *Anal. Chem.* 67, 1153–1158
13 Kiernan, U.A. *et al.* (2004) Proteomic characterization of novel serum amyloid P component variants from human plasma and urine. *Proteomics* 4, 1825–1829
14 Nedelkov, D. and Nelson, R.W. (2003) Detection of staphylococcal enterotoxin B via biomolecular interaction analysis mass spectrometry. *Appl. Environ. Microbiol.* 69, 5212–5215
15 Kiernan, U.A. *et al.* (2003) Detection of novel truncated forms of human serum amyloid A protein in human plasma. *FEBS Lett.* 537, 166–170
16 Nedelkov, D. and Nelson, R.W. (2003) Surface plasmon resonance mass spectrometry: recent progress and outlooks. *Trends Biotechnol.* 21, 301–305
17 Nedelkov, D. *et al.* (2003) Detection of bound and free IGF-1 and IGF-2 in human plasma via biomolecular interaction analysis mass spectrometry. *FEBS Lett.* 536, 130–134
18 Won, Y. *et al.* (2003) Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. *Proteomics* 3, 2310–2316
19 Zhang, R. *et al.* (2002) Controlling deuterium isotope effects in comparative proteomics. *Anal. Chem.* 74, 3662–3669
20 Wehr, T. (2003) Affinity selection techniques for proteomic studies. LC/GC 21, 274-84
21 Abrams, S.A. and Wong, W.W. (2003) *Stable Isotopes in Human Nutrition: Laboratory Methods and Research Applications*, CABI Publishing
22 Schoenheimer, R. and Rittenberg, D. (1935) Deuterium as an indicator in the study of intermediary metabolism. *Science* 82, 156–157
23 Hellerstein, M.K. Method for measuring *in vivo* synthesis of biopolymers. US Patent 5,338,686 (Apr. 16, 1994).
24 Ong, S.E. *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386
25 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
26 Rose, K. *et al.* (1983) A new mass-spectrometric C-terminal sequencing technique finds a similarity between γ-interferon and α2-interferon and identifies a proteolytically clipped γ-interferon that retains full antiviral activity. *Biochem. J.* 215, 273–277
27 Prolytica™ [^{18}O] Labeling Kit: Instruction Manual, Catalog # 271010 Rev. # 054001 (Stratagene, La Jolla, CA, 2004).
28 Mirgorodskaya, O.A. *et al.* (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using ^{18}O-labeled internal standards. *Rapid Comm. Mass Spectrom.* 14, 1226–1232
29 Stewart, I.I. *et al.* (2001) 18O labeling: A tool for proteomics. *Rapid Commun. Mass Spectrom.* 15, 2456–2465
30 Chakraborty, A. and Regnier, F.E. (2002) Global internal standard technology for comparative proteomics. *J. Chromatogr. A.* 949, 173–178
31 Regnier, F.E. *et al.* Affinity selected signature peptides for protein identification and quantification. WO0186306 (Nov. 15, 2001).
32 Liu, P. and Regnier, F.E. (2002) An isotope coding strategy for proteomics involving both amine and carboxy group labeling. *J. Proteome Res.* 1, 443–450
33 Zhang, R. and Regnier, F.E. (2002) Minimizing resolution of isotopically coded peptides in comparative proteomics. *J. Proteome Res.* 1, 139–147
34 Regnier, F.E. and Zhang, R. Controlling isotope effects during fractionation of analytes. WO03027682 (Apr. 3, 2003).
35 Zhang, K. *et al.* New technologies for expanding the dynamic range of protein identification in human serum. Poster presented at Am. Soc. Mass Spectrom. 52nd Annual Meeting. Nashville, TN (May 24, 2004).
36 Peters, E.C. and Horn, D. (2001) A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun. Mass Spectrom.* 15, 2387–2392
37 Pappin, D.J.C. and Bartlet-Jones, M. Methods, mixtures, kits and compositions pertaining to analyte determination. WO 2004/070352 (Aug. 19, 2004).
38 Ren, D. *et al.* (2004) Histidine-rich peptide selection and quantification in targeted proteomics. *J. Proteome Res.* 3, 37–45
39 Wang, S. *et al.* (2002) Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. *J. Chromatogr. A.* 949, 153–162
40 Aebersold, R.H. *et al.* Rapid quantitative analysis of proteins or protein function in complex mixtures. US6670194 (Dec. 30, 2003).
41 Han, D.K. *et al.* (2001) Quanitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 19, 946–951
42 von Heller, P.D. *et al.* (2003) The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation. *Mol. Cell. Proteomics* 2, 428–442
43 Hansen, K.C. *et al.* (2003) Mass spectrometric analysis of protein mixtures at low levels using cleavable ^{13}C-isotope-coded affinity tag and multidimensional chromatography. *Mol. Cell. Proteomics* 2, 299–314
44 Li, J. *et al.* (2003) Protein profiling with cleavable isotope coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol. Cell. Proteomics* 2, 1198–1204
45 Zhang, H. *et al.* (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* 8, 66–75
46 Anderson, N.L. and Anderson, N.G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867
47 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).
48 Schneider, L.V. *et al.* Mass defect labeling for the determination of oligomer sequences. US20020172961 (Nov. 21, 2002).
49 Schneider, L.V. *et al.* (2004) Mass-defect tagging for proteomic analysis. *Genetic Eng. News* 24, 28–30
50 Hall, M.P. and Schneider, L.V. (2004) Isotope differentiated binding energy shift tags (IDBEST™) for improved targeted biomarker discovery and validation. *Expert Rev. Proteomics* 1, 1–11
51 Hall, M.P. *et al.* (2003) 'Mass defect' tags for biomolecular mass spectrometry. *J. Mass Spectrom.* 38, 809–816
52 Whetstone, P.A. *et al.* (2004) Element-coded affinity tags for peptides and proteins. *Bioconjug. Chem.* 15, 3–6
53 Hernandez, H. *et al.* Improvements in mass defect labeling for shotgun proteomic analysis. Poster presented at the 52nd Ann. Mtg., *Am. Soc. Mass Spectrom.*, Nashville, TN (May 23-26, 2004).
54 Yates, J.R., III and Eng, J.K. Use of mass spectrometry fragmentation patterns of peptides to identify amino acid sequences in databases. US5538897 (July 23, 1996).
55 Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66, 4390–4399
56 Kelleher, N.L. (2004) Top-down proteomics. *Anal. Chem.* 76, 197A–203A
57 McLafferty, F.W. *et al.* (1998) Two-dimensional mass spectrometry of biomolecules at the subfemtomole level. *Curr. Opin. Chem. Biol.* 2, 571–578
58 Schneider, L.V. *et al.* Methods for sequencing proteins. US6379971 (Apr. 30, 2002).
59 Krause, M. *et al.* Method and reagent for specifically identifying and quantifying one or more proteins in a sample, WO04001420 (Dec. 31, 2003).