Amino Acids (2013) 45:219–229
DOI 10.1007/s00726-013-1505-1

REVIEW ARTICLE

Combinatorial peptide libraries to overcome the classical affinity-enrichment methods in proteomics

Pier Giorgio Righetti · Egisto Boschetti

Received: 18 April 2013 / Accepted: 20 April 2013 / Published online: 30 April 2013

Abstract The enrichment of targeted low-abundance proteins is possible by affinity adsorption using selected sorbents. Different categories of very dilute proteins are present in most of biological extracts so that specific affinity methods are unable to address their collective enrichment. Only recently an interesting approach has been proposed associating the affinity of multiple ligands used as a library mode under overloading much beyond the saturation of the affinity mixed bed. The principle and the limits of this technology are reported along with their current and potential applications in various domains.

Keywords Enrichment · Proteomics · Low-abundance proteins · Peptide library · Marker discovery

Brief introduction

The complexity of proteomes with hundreds of thousands of polypeptides and a dynamic concentration range of proteins that extends over 12 orders of magnitudes clearly suggest that the majority of species are of low- or very low-abundance. In the majority of the cases it is postulated that they are below the detection sensitivity of current analytical methods. In this situation, it is not strange to see that many scientists are delving in depth in this category of proteins because they may represent very important functions at a given stage of the cell cycle or the cell maturation. The discovery of these entities should effectively contribute to elucidate unknown pathways and also to understand the rules of activation or inhibition of expression mechanisms.

The investigation of all these species necessitates their selective concentration or even their isolation prior to analysis of both their structure and function.

Since the beginning of proteomics investigations, scientists implemented methods for this purpose by not only optimizing the existing separation methods, but also by devising novel approaches capable to act globally toward the low-abundance species or to target specific proteins or selected families. It is here recalled that extensive fractionation protocols have been implemented for discovering minor proteins in human blood plasma (Faca et al. 2007). Other approaches were based on the use of narrow- or very narrow-range pH gradients in isoelectric focusing prior to the second dimensional migration of proteins within the two-dimensional electrophoresis space (Westbrook et al. 2001). Some abundant proteins would therefore be excluded from the separation gradient and, with massive protein loadings; a large number of proteins could become detectable as new spots. However, at loads comprising the entire protein extract, massive precipitation is generated just outside the selected pH gradient entrapping thus many low-abundance species that do not migrate and thus escape detection.

The detection of low-abundance species is also prevented because of the presence of a few extremely concentrated proteins such as albumin in serum. To resolve this issue the specific removal of certain proteins has been successfully proposed. Initially, this was limited to albumin and IgGs but rapidly this so-called "depletion" process was

P. G. Righetti
Miles Gloriosus Academy, Via Archimede 114, 20129 Milan, Italy

E. Boschetti (✉)
Scientific Consultant, Paris, France
e-mail: egisto.boschetti@gmail.com

extended to 6 (Gong et al. 2006), 12 (Huang and Fang 2008), 20 (Levin et al. 2007) and more proteins (Gao et al. 2008). Unfortunately, these approaches are of poor efficacy because they contribute to further dilute rare species obliging the operators to concentrate the resulting solutions with significant protein losses. This drawback adds to the co-adsorption of low-abundance species (Bellei et al. 2011) negating their detectability and much reducing the interest of such an approach. Thus, one is forced to consider that only affinity-based methods could resolve the question of enriching for very low-abundance species. Using this technology effectively a real enrichment can be achieved. Many examples are described in the literature and some are given below. Nevertheless, all these approaches are very selective for single proteins or groups of proteins.

In these last years a novel approach still based on affinity interactions has been proposed for the enrichment of general low-abundance species without discrimination of shape, isoelectric point, size, composition or function. This is the use of affinity ligands grafted on beads that are then mixed together and used as a mixed bed under a large overloading. Next sections will discuss these approaches and make a synthesis of what is known to date.

Single-feature affinity-enrichment approaches

By single feature, it is here meant that the affinity-based enrichment is obtained using a single ligand grafted on a solid support. The targeted species are indistinctly single proteins or groups of proteins sharing similar composition or similar functions. Although a quite common belief is that current fractionation methods, such as cation exchange chromatography (Edelmann 2011), could concur to an enrichment effect, it is here considered preferable to adopt affinity or pseudoaffinity interactions.

Most of the time the solid phase (e.g., chromatographic sorbent) comprises beads grafted with a single ligand of narrow specificity used for instance to group proteins of common post-translational modifications. However, there are also proteins that have innate affinity for common ligands such as nucleic acid-binding proteins or proteins with similar metal ion affinity. The use of such affinity media contributes to enhance the concentration of preselected species up to the detectability level using common analytical methods.

Perhaps the most prominent example is the use of antibodies addressing a given protein antigen. Beyond single proteins that can easily be enriched using grafted antibodies, a common route of enrichment is the use of ligands targeted to groups of proteins.

Heparin for example is a well-known affinity ligand used for the purification of numerous proteins. It naturally interacts with coagulation factors (Clifton et al. 2010) and with growth factors (Hnasko and Ben-Jonathan 2003). It displays an affinity activity for many other proteins and was consequently used in a number of applications to enhance the concentration of low-abundance proteins. This is the case of low-abundance tissue-specific biomarkers found in pulmonary adenocarcinoma (Rho et al. 2009) and for the detection of low-abundance signaling proteins with the aim of finding drug targets (Krapfenbauer and Fountoulakis 2009).

In spite of some technical limitations, lectins, polypeptide ligands derived mostly (but not solely) from plants, are also extensively used for glycoprotein enrichment with the double objective of: (1) discovering rare species and (2) segregating special posttranslational modifications resulting from biological misregulations. The most popular lectins used for the separation of glycoproteins are concanavalin A (ConA), wheat germ agglutinin (WGA), jacalin and *Ulex europeus* agglutinin (UEA) allowing a good coverage of the panel of glycoproteins existing in eukaryotic cells. As reported by Drake et al. (2006), to obtain a complete glycoprotein coverage a set of six different lectins is sufficient. Due to their diverse specificity, they are generally used singularly or as a sequential chromatography cascade (Yang et al. 2006). Tracking differences in the glycan structure of certain proteins could represent an interesting option for diagnostic purposes. In cancer conditions, alterations of the glycan at the fucosylation and sialylation points are demonstrated. In other circumstances changes in the glycan branching are also frequently observed. The significance of structural "errors" within low-abundance glycoproteins could be considered as a first sign of an upcoming or an early-stage disease. For references see following reviews: An et al. (2009); Tian and Zhang (2010); Butterfield and Owen (2011); and Adamczyk et al. (2012).

Nucleic acids are known to bind with a number of low-abundance proteins, hence they are considered as ligands for the enrichment of corresponding binding proteins. They certainly deserve attention when dealing with the elucidation of certain biological disorders or with the clarification of nuclear biological pathways where very low-abundance proteins are involved. PolyU has been used as a ligand for the detection of RNA-binding proteins from *Arabidopsis thaliana* leaf extracts that generally escape the current two-dimensional electrophoresis analysis (Xu et al. 2007). In this study the authors identified species expressed at levels lower than 0.1 %. Immobilized PolyA has also been used to identify rare chloroplast proteins expressed at levels that cannot be detected directly without enrichment processes (Ni et al. 2010).

Similarly, cyclic nucleotide binding proteins of low-abundance, involved in signaling pathways, are very good

targets for performing in-depth investigations. These ligands cannot be genuine cyclic nucleotides because they are not directly graftable on solid supports, thus they are chemically modified prior to immobilization on beads (Medvedev et al. 2012). The elucidation of the specific interaction between phosphokinase-A and A-kinase anchor proteins (PKA-AKAP) has been accomplished after the enrichment of PKA by means of three different cyclic AMP analogs attached to agarose beads. The grafting has been possible by the introduction of an aminohexylamino spacer in two different positions or by the introduction of another chemical group at the position 2′ of the ribose. More than ten AKAPs have been identified, each of them possibly involved in different PKA isoforms activation in various cell compartments. Other cyclic nucleotide derivatives such as cGMP have been chemically attached on agarose beads to enrich for interacting proteins (Scholten et al. 2006). Beside known interacting proteins, such as kinases and phosphodiesterases, other proteins have been co-captured including several AKAP. The reduction of non-specifically co-adsorbed species has been achieved by various washings with selective displacement agents such as ADP and GDP.

Metal ion-mediated interaction with proteins is another way to enrich for pre-determined species. This technology is based on the ability of mildly chelated metal ions to interact with some well-known exposed peptide side chains. The most known chelating agents used are iminodiacetic acid and nitrilotriacetic acid. Although frequently the protein chemical group involved is a histidyl residue, other amino acids such as tryptophan, phenylalanine and arginine can also contribute to the interaction phenomenon (Porath and Olin 1983). Once metal ion binding proteins are captured, they can be harvested by a simple elution involving a reduction of pH or the use of competition agents such as imidazole or chelating agents.

Different groups of proteins can be targeted depending on the nature of the chelated metal ion: phosphoproteins are preferably enriched using trivalent iron ions (Li and Dass 1999), gallium ions (Posewitz and Tempst 1999) and other transition species such as aluminum and cobalt ions. Since other non-phosphorylated proteins can be captured by their exposed histidine side chains, the technology can be improved by the acidification of the separation conditions. Many new diverse phosphorylated species have been identified after trypsinization and sophisticated mass spectrometry procedures. When the number of phosphorylated structures present is very large [it is estimated that half of the proteins can be phosphorylated with up to 100,000 phosphorylated sites in a human proteome (Zhang et al. 2002)], phosphopeptides are fractionated prior to mass spectrometry analysis with a significantly increased potential discovery (Stone et al. 2011).

With this affinity technology different metal ions have been described. An interesting example is given by the separation of uranyl-binding proteins obtained after having chelated uranyl cations on a solid-phase resin (Basset et al. 2008). The purpose of the study was to enhance a sub-proteome with respect to radiation effects. On other examples enrichment processes for metallo-proteomes have been reported from Dead Sea archaeal cells (Kirkland et al. 2008) or from marine cyanobacteria with properties to bind cobalt, iron, manganese and nickel ions (Barnett et al. 2012).

Other more or less complicated single-feature affinity-enrichment approaches are applied to various types of proteins, protein categories and sub-proteomes. They are for instance specific antibodies for the capture or rare proteins such as phosphorylated species where the phosphoric ester is located either at the serine or at the threonine levels (Gronborg et al. 2002). Ubiquitinylated protein groups are also enriched using immunoaffinity chromatography (Shimada et al. 2008). Carbonylated proteins, a category related to the age, are enriched for in-depth analysis after proper dinitrophenyl-hydrazide tagging and then captured by immobilized antibodies against dinitrophenyl-hydrazide (Madian et al. 2011).

Multiple-feature affinity-enrichment methods

As a corollary of the previous paragraph, multiple-feature affinity capture is another way to enrich proteins that are of low- or very low-abundance in current biological fluids or extracts. By “multiple-feature” the concomitant action of two or more solid-phase affinity media mixed in a single bed is intended. This is generally obtained by simply mixing different affinity resins and using the resulting blend for the capture and hence the simultaneous enrichment of various proteins. A mixed bed of distinct sorbents necessitates a good compatibility of at least adsorption conditions at which all proteins are submitted at the same time. This is the reason why blending sorbents is not very common in liquid chromatography. Nonetheless, this approach has been described for most popular chromatographic supports such as ion exchangers and hydrophobic solid phases (Maa et al. 1998). In another example, an ion exchange mixed bed column was successfully described for the separation of phosphopeptides from enzymatically digested proteins (Motoyama et al. 2007). More recently, the use of ion exchange mixed bed proved enhanced capability to enrich for a secreted *O*-glycoproteome (Darula et al. 2012).

In proteomics the classical example of mixed bed chromatography application is the simultaneous subtraction of most abundant serum proteins using a mixed-mode

immunodepletion. Rather than enriching for given species, this type of technology is intended to remove several highly abundant proteins (e.g., albumin, IgG and others) that generally prevent the detectability of minor components (Pieper et al. 2003). Several antigen-specific affinity immunosorbents are first made and then mixed together before use. Each immunosorbent is compatible with a physiological buffer for protein capture and can consequently be used as a mixed bed. In spite of this interesting approach that eliminates the burden of most abundant proteins interfering with the detectability of lower concentration species, the process does not enrich for low-abundance species, those categories that are below the detectability level of analytical methods (see Sect. "Introduction").

Lectins—ligands for glycoproteins—are also used immobilized as mixed bed for general adsorption of a large spectrum of glycoconjugates. They have been used as mixed bed for the analysis of glycoproteomes with good reproducibility (Kullolli et al. 2008). Semiquantitative data have been produced by spectral counting to determine abundance differences among clinical samples. In a more sophisticated approach two mixed bed columns could be used: the first for immunodepletion of abundant species and the second for the enrichment of glycoproteins.

Combinatorial solid-phase affinity-enrichment

Perhaps the best example of a mixed bed for the enrichment of proteins that are all considered of low abundance is the use of solid-state combinatorial ligands where the obvious representative is a peptide library. From their random composition these peptides display a sort of continuum in the affinity strength for proteins. Small peptide differences influence also protein displacements effects due not only to the affinity differences, but also to their respective concentrations, in virtue of the mass action law. These phenomena have extensively been described in a book dedicated to the use of peptide library for the analysis of low-abundance proteins (Boschetti and Righetti 2013).

The principle

The first report describing a general enrichment process of proteins undetectable with regular technologies was published by Thulasiraman et al. (2005). It addressed the situation of proteins of very low concentration independently from their structure or function. For the first time the concept of using overloading conditions to reach the goal was adopted. This approach allowed also reducing the concentration of high-abundance proteins resulting into the notion of "reduction of dynamic concentration range". From this seminal paper, many scientific reports followed as well as general reviews. All the principles, the molecular mechanisms and the applications are assembled in a book dealing with low-abundance proteins where many detailed protocols are also included (Boschetti and Righetti 2013, pp: 262–314). The technology uses mixed beds of chromatography beads that are all different from each other since individually they carry different affinity ligands. Each bead, considered as a single affinity solid phase, addresses a single or a group of proteins having in common the capability to dock on the same peptide ligand and theoretically having similar binding capacity. Due to the very numerous peptide ligands (several millions), most if not all proteins are targeted. Under large overloading conditions, very dilute species are concentrated on their respective beads while concentrated (high-abundance) species saturate very rapidly their bead partners and remain in the supernatant from where they are washed away. In spite of a common initial belief that each bead would address a given protein, the reality is more complex. There are different proteins that have an affinity for the same peptide ligand with two important differences that are their affinity constants and their relative concentrations. This situation induces an intense competition all along the loading phase until equilibrium. This is reinforced by the presence of very similar peptide ligands that might differ from each other by just one amino acid. The adsorption of proteins is therefore the result of a sort of equilibrium of contradictory forces generated by the mixed-mode ligand (the peptide) that contains electrical charges acting synergistically or antagonistically to the docking phenomenon. Hydrophobic associations and hydrogen bonding are promoted by other amino acids. Globally, attraction and repulsion forces generated by the mixed-mode nature of the peptide ligand cover the full spectrum of molecular interactions as described in detail (Boschetti and Righetti 2013, pp: 99–112).

After having washed out the excess of proteins, those that have saturated their respective beads—the captured species—are desorbed and collected by various methods depending on the type of analysis.

This technology is modulated by a relatively large spectrum of physicochemical parameters such as the number of diverse peptides, which depends on the number of amino acids used and the length of the final peptide, the particle size of the beads, the conditions of protein capture (extent of overloading, pH and ionic strength of the buffer, temperature, protein concentration, presence of competitors, etc.) and the desorption method.

Although very powerful to enrich for very dilute proteins, the described technology suffers from a few limitations intimately related to the concept itself. First, the amount of sample engaged in the process can be considered

quite large but it is common to all enrichment methods. The relatively large sample required is actually a distinct advantage of enrichment since with very small biological samples (e.g., 10 µL serum), one would never have a chance of finding protein traces that might comprise possible early-stage biomarkers present at pico- or sub-pico grams/mL level. The second limitation is that some proteins of medium- or high-abundance currently found without the treatment of the sample are undetectable after the contact with the peptide library. This issue is in reality reduced close to zero if (1) different pHs and different ionic strengths are used for the capture phase and if (2) the stringency of elution process is strong enough to collected species that are very tightly adsorbed. While weakening certain interaction forces capable to desorb some species other forces can easily be exacerbated, so that the protein desorption would not be complete. It is with these questions in mind that a series of protocols, some being absolute desorption solutions and other sequential solutions, have been proposed. The selection is generally performed according to the following analytical determinations. It is interesting to recall that sodium dodecyl sulfate—SDS (e.g., Laemmli buffer)—is probably the most effective desorption agent. The collected proteins can thus directly be analyzed by SDS-PAGE followed by gel slicing, trypsinization and LC–MS/MS. However, if the analytical determination is not compatible with the presence of SDS (e.g., 2D-PAGE, MALDI, etc.) the protein solution must be cleaned up with protocols that are detailed in Boschetti and Righetti (2013), pp: 273–276. Alternatively, the proteins captured by the library beads can be analyzed without elution. This is possible by direct proteolysis on beads followed by peptide fractionation and mass spectrometry (Fonslow et al. 2011).

In many cases, the findings from the treatment of biological extracts with solid-phase combinatorial peptide libraries are really extraordinary with a large increment of the number of proteins detected that would not otherwise be visible. Figure 1 represents selected examples from many reported data.

Necessity of overloading conditions

The enrichment of low-abundance proteins subsequent to combinatorial peptide library treatment is the result of a large bead overloading. The saturation loading is about 3 ng of protein per bead toward the corresponding protein partner; therefore, it is useful to approach this level if one aims at best enrichment and highest detection of very dilute species. In a given sample the large number of proteins present covers a very large range of concentrations and when targeting for the low-abundance ones the sample volume should be large enough to allow enrichment of

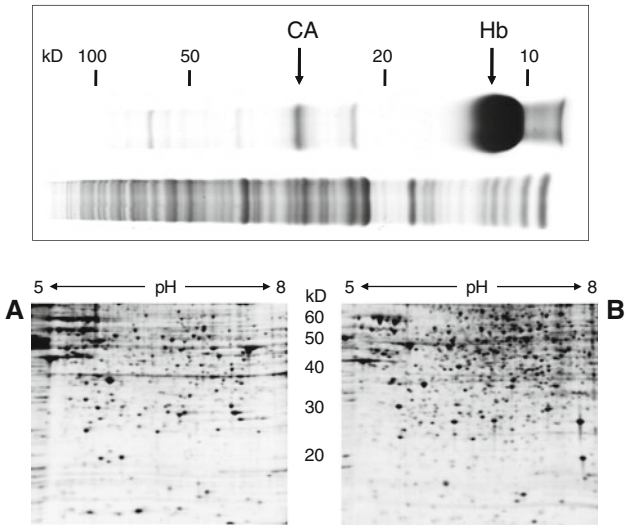

Fig. 1 Examples of low-abundance proteins enrichment using a solid-phase peptide library. The *upper panel* represents the treatment of red blood cell proteins extract before (**a**) and after (**b**) treatment. An increased number of protein bands are visible along with a strong decrease of the major band of hemoglobin (Hb). *CA* carbonic anhydrase. Unpublished data from C. Simò and P. G. Righetti. *Lower panel* two-dimensional electrophoresis of HeLa cell lysates before (**a**) and after (**b**) treatment with a solid-phase peptide library. Narrow focusing gradient between pH 5 and 8. Data from Fonslow et al. (2011) by kind permission

low-abundance proteins. This means that to get enough proteins present in traces, the initial sample volume could be significant. Naturally, most concentrated species will be in large excess, which is not detrimental to the process because the supernatant where they are present is eliminated continuously (use of combinatorial beads in column) or after incubation (use of combinatorial beads in a small batch).

Demonstration of this phenomenon has been performed repeatedly. Figure 2 illustrates the results for different loading conditions. Very simply, when the amount of protein is increased the beads become progressively saturated. Below bead saturation the compression of the dynamic range is low or even irrelevant depending on the initial concentration difference between high and low-abundance proteins. In many cases, large sample volumes contributed to the discovery of novel species never described before. This has been observed experimentally by Roux-Dalvai et al. (2008). These authors identified 1,288 gene products from red blood cell lysate when the loading was about 5,700 mg of proteins per 1 mL of combinatorial beads compared to 331 gene products found in the non-treated sample for the same amount of proteins analyzed. However, if the initial bead loading is lower the number of gene products found is also smaller. Rare ε and ζ embryonic hemoglobin chains supposed to be repressed in mature

cells have been found with such a large load; however, they were undetectable when treating small amounts of cell lysates.

While attempting the enrichment of glycoproteins of low abundance, Huhn et al. (2011) demonstrated that at very large loading conditions additional protein bands appeared upon SDS-PAGE. Experiments were performed using 250, 500, 750 and 1,000 μL of human plasma mixed with 100 μL beads. A demonstration was made that protein patterns are enriched of additional new bands as the load increased.

The enrichment phenomenon could take a different aspect according to the composition of the sample. With protein extracts from chicken pectoral muscles, Rivers et al. (2011) effectively found the enhancement of new gene products; however, they also observed the formation of novel high-abundance species when the amount of exposed proteins to the beads was very large. Here, also the experiments were organized around the same volume of beads (100 μL) loaded with 20, 50, 100, 250, 500 and 1,000 mg of proteins. Numerous species were newly detected at a level that was impossible to find prior to sample treatment. Among other data the authors found that the skeletal muscle was essentially dominated by glycolytic enzymes involved in muscle contraction with a profound change from the untreated sample. A significant dynamic protein concentration range compression was produced with a shallower protein abundance distribution, on the one hand, and a correlation between the appearance of low-abundance species and the sample load increasing, on the other hand.

The examples reported demonstrate the capability of this technology to probe rare proteins when the loading is large improving, thus, the understanding of proteome compositions and suggesting to facilitate the biomarker discovery at early stages of a disease.

The requirement of single ligands carried by individual beads

In spite of theoretical explanations of the mode of action of ligand libraries to enrich for very dilute species, various experimental demonstrations against a proper control have been organized. They contributed to resolve wrong interpretations not only on the composition of the ligands but also on the immobilization topography. In a published paper of 3 years ago, Keidel et al. (2010) stated that the peptide libraries work according to a regular hydrophobic association of a reverse phase adsorption and that the diversity of peptides has only a marginal effect. In reality, this statement that goes against the described mechanism did not resist to experimental evidence. The molecular mechanisms of a library (mixed bed of mixed-mode ligands) contrast considerably from a homogeneous bed of a single-mode ligand (e.g., hydrophobic sorbent) because of (1) the interactions mechanisms involved are multiple and concomitant and not singular (electrostatic, hydrogen bonding, hydrophobic, van des Waals) as extensively described (Boschetti and Righetti 2013), (2) each bead acts independently from another one both having different peptide ligands, (3) salts such as sodium chloride desorb a large part of captured proteins meaning that hydrophobic interaction is not the dominant capture mode (D'Amato et al. 2010) and (4) it has been repeatedly demonstrated that it was possible to fish out from the peptide library bead specimens that are specific for a given protein and not for another (Lathrop et al. 2007), this operation being clearly impossible with the use of either a singular hydrophobic solid phase or even an ion exchanger. To all the above, it is to be added that more proteins are captured by a peptide library mixed bed when the salt concentration is diminished (Di Girolamo et al. 2011) and that it has been largely documented since years that hydrophobic interaction does not occur at low ionic strengths.

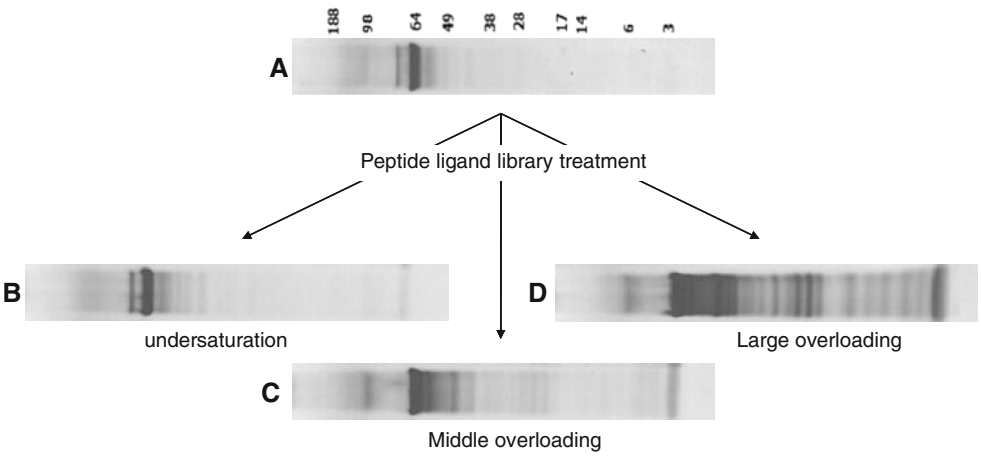

Fig. 2 SDS-PAGE analysis of lymphocyte cell culture supernatant **a** treated with combinatorial peptide library. **b–d** Eluates obtained from various protein loadings with respective bead to sample volume ratio of 1:1, 1:10 and 1:100, adapted from Thulasiraman et al. (2005)

Contrary to a hydrophobic resin for protein binding, a library is a collection of a large number of solid phase media mixed together, each bead carrying a different grafted chemical structure. Here the real legitimate question would be whether the peptide ligand would need to be grafted separately (one-bead-one-peptide) or grafted all together on all beads (one-bead-all-peptides).

To demonstrate the different behavior of these two possibilities specific experiments have been performed. Two different grafted peptide libraries have been made in parallel. These ligands were hexapeptides issued from the same synthesis with the difference that the first library was composed of a mixed bed of one-bead-one-peptide and the other was composed of a homogeneous bed of one-bead-all-peptides. These two sorbent beds have been overloaded with an equal large excess of human serum and after having washed away the excess of proteins, the captured species were globally desorbed and analyzed in parallel by two-dimensional electrophoresis. Figure 3 shows the difference in protein patterns. In both cases, the level of abundant proteins as for instance albumin was largely decreased; however, the one-bead-one-peptide mixed bed clearly adsorbed a large number of proteins of low abundance, while the one-bead-all-peptide resin was unable to capture as many proteins. In addition the latter generated at least one abundant species identified as apolipoprotein A1. The dissimilar protein pattern is explained by the different mechanism of protein competition described in Sect. "Necessity of overloading conditions". In a bead comprising a single peptide structure the competition is resolved by both the difference of affinity for the same ligand and the concentration difference between competitors. In the second case (a bead comprises all peptide structures) a protein interacts with several different peptides at the same time in accordance with compatible atomic distances. The competition for a single peptide ligand is here almost non-existing. Therefore, the saturation phenomenon due to overloading conditions appears only as a secondary effect on the enhancement of low-abundance proteins, because a protein that docks under different interacting points is very difficult to displace by another protein that may not have the capability to interact with the same set of peptides. The outcome is that the final composition of eluted proteins is largely different from what results from the one-peptide-one-bead library (Boschetti and Righetti 2008). As a consequence, it appears that the one-bead-all-peptides situation does not seem giving advantages to the capture of low-abundance species because of the poor possibility of displacement. This is why it reveals less numerous spots in two-dimensional electrophoresis and creates new high-abundance proteins capable to interact simultaneously with several peptide structures placed close to each other within the same bead. In this situation proteins that have natural good propensity to interact with a large number of surfaces (e.g., apolipoproteins A1) have the advantage over other proteins.

Applicability of combinatorial peptide library in protein enrichment

If one is looking for an enrichment method for a single protein or a group of proteins, the use of combinatorial ligands libraries is the wrong approach. Such a method addresses the general enrichment of all species independently from their shape, composition and function, while concomitantly reducing the concentration of high-abundance ones.

Throughout the numerous papers published the applicability of the described technology is very large, encompassing the elucidation of the composition of proteomes with respect to undetectable polypeptides, the discovery of early-stage biomarkers, the detection of novel rare allergens and the search of protein traces as contaminants in food, beverages and pure biopharmaceuticals.

This mini-review does unfortunately not allow detailing each of these applications, therefore, only a few examples will be given. The readers are invited to find application details and protocols of use in a dedicated book (Boschetti and Righetti 2013) and in other major reviews (Righetti et al. 2010a, b; Boschetti and Righetti 2012).

Clearly one of the major applications in protein enrichment is to find protein markers of diagnostic or prognostic interest from various biological materials such as circulating fluids or tissue extracts. While the route is long and tedious due to numerous controls, very large sample cohorts, validation against other possible overlapping markers related to other diseases, the technical approach is very compelling as recently described (Boschetti and Righetti 2012). The chances of finding relevant protein markers can be enhanced significantly if the described technology is associated with one or more adapted analytical technologies (Su et al. 2013).

Associated to two-dimensional electrophoresis (Boschetti et al. 2013) it becomes a very powerful tool to discover new low-abundance allergens of medical interest when the reactions against pre-determined aggressive agents (e.g., pollens) are negatives with respect to standard allergens (Shahali et al. 2012). This is a very large field of investigation for at least two major reasons: (1) the codified allergens are of medium or high abundance and (2) more and more frequent allergic reactions are reported and are increasingly aggressive. This phenomenon is not only related to seasonal pollens, but more importantly it originates from food and food components. This is the reason why food science is taking momentum. In this respect one has to consider that food (especially industrial food) can be

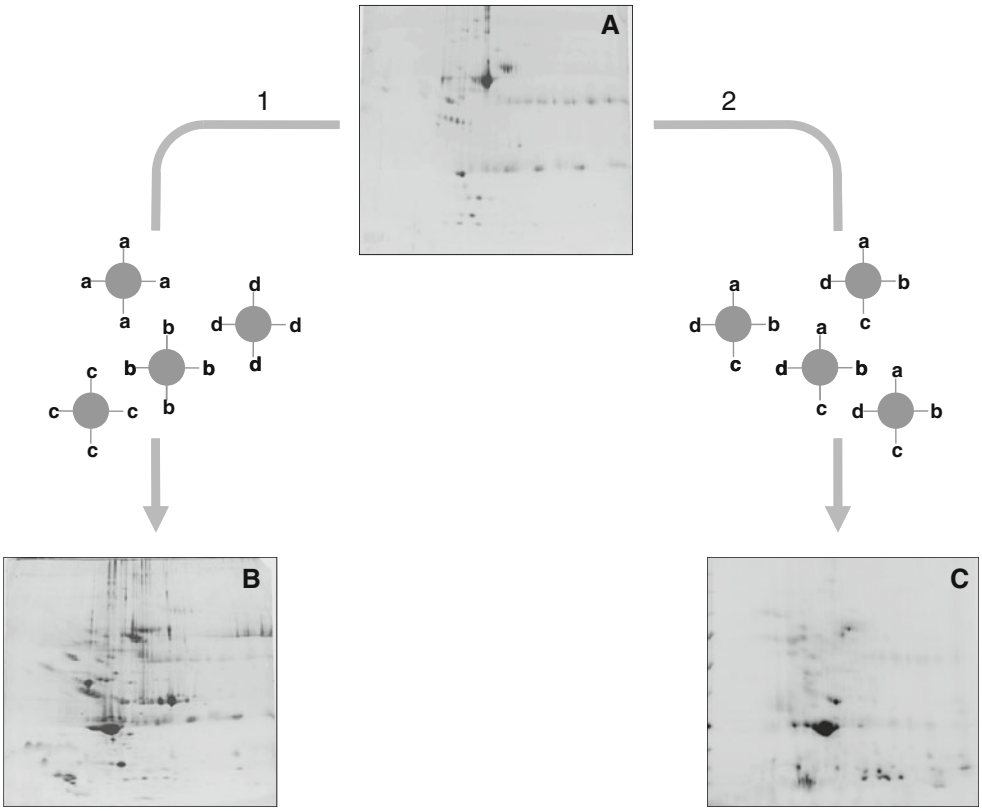

Fig. 3 Effect of grafted peptide library ligand topography on the ability to capture protein species. **a** Two-dimensional electrophoresis of the initial untreated sample of human serum. **b** Two-dimensional electrophoresis of proteins captured by peptides grafted on beads as one-bead-one-peptide. **c** Two-dimensional electrophoresis of proteins captured by peptides grafted on beads as one-bead-all-peptides. pH focusing gradient between 3 and 10, staining with Coomassie blue, **a**–**d** schematically represent the diverse peptides. The library consisted in hexapeptides made with 16 major amino acids

of very complex composition with ingredients of animal and vegetal origin possibly comprising a large number of low-abundance allergic species. As described recently, it is anticipated that the use of combinatorial peptide library technology allowing the "amplification" of very dilute species will significantly contribute to the discovery of incompatible ingredients (Righetti and Boschetti 2012). This approach has already been extended to the search of components that are not always allowed to be present at least above certain concentrations. This is the case of the presence of proteins in wine that only find justification as stabilizing agents such as caseins and egg white proteins. Both are tolerated during wine processing providing their presence is undetectable in the final products as a demonstration that all traces are eliminated. With the described technology, it has been found that these proteins are still present in various commercial wines (Righetti et al. 2012) justifying the extension of this investigation to other beverages as well.

Biopharmaceuticals too may comprise contaminants of proteinaceous origin from host cell or from culture medium with numerous side effects for patients especially when the treatment is prolonged and massive. Purity of biopharmaceutical preparations meant for human consumption is, however, an unclear notion, because the presence of detected (or undetected) impurities depends on the analytical determinations. The use of peptide libraries can easily contribute to find protein traces and, in our opinion, it should be generalized to enrich for general (not targeted) contaminants with the object of detecting species that could be aggressive for the receiving patient. Once the presence is demonstrated it will be easy to remove them with the consequent improved safety of the final product (Righetti et al. 2011).

At this stage one would ask the question of the quantitation aspect of these low-abundance species since the initial sample is "manipulated" and would not correspond any more to the real initial sample. This argument is in reality not receivable for reasons largely debated and explained in various papers and also detailed in Boschetti and Righetti (2013). Simply most of the time the detectability needs to be dissociated from the enrichment: the latter serves the discovery stage and once this is demonstrated, the quantitation can easily be assessed with immunoassays of various nature and good enough sensitivity.

Conclusions and future potential developments

The enrichment of predetermined gene products or of protein groups is today possible by means of selective solid phases and more particularly by affinity resins. The enrichment process depends on the amount of initial proteins able to bring enough material when the targeted species are

present at extremely low concentration. A number of strategies have been devised to prevent non-specific binding by stringent selection of ligands or by preliminary modification of targeted groups of proteins or by specific displacement effects after the first capture stage. All these approaches are derived from chromatographic technologies developed a couple of decades ago. However, the question of low-abundance proteins enrichment taken globally cannot be resolved by the above mentioned approaches because they are at most groupable within a predetermined functional or structural group. Interestingly, classical affinity chromatography brought a number of ligand concepts that could be exploited for peculiar situations. Mixed beds of combinatorial affinity ligands able to adsorb at the same time an extremely large number of proteins present in a sample (if not all proteins present) is one of the possible solutions. When these systems are used under overloading conditions superimposed to the affinity phenomenon, the low-abundance species become singularly concentrated and easily analyzable by current methods. This is an extremely strong tool at the disposal of biochemists that are concerned by low- or very low-abundance protein discovery.

Development avenues opened by these affinity-based methods are numerous such as finding heterologous protein traces where they are supposed to be absent (e.g., wines), early-stage biomarkers of metabolic alterations, rare but powerful allergens discovery, just to mention a few.

Generic low-abundance protein enrichment can easily be followed by a more specific affinity-enrichment process to sub-fractionate categories of proteins of similar biological function or similar structure (e.g., low-abundance phosphoproteins, low-abundance glycoforms, etc.).

It is also conceivable to design special libraries targeting proteins by their general properties (outside their function and their structural similarity) such as hydrophobic species or cationic low-abundance species or extreme isoelectric point low-abundance proteins. However, what should be kept in mind is that all approaches involving mixed beds should be used under large overloading conditions followed by a complete desorption of all species prior to analysis. Needless to say, when such stringent rules are not followed, disasters occur, as in the case of Bandow (2010), who could not find any low-abundance proteins in sera treated with ligand libraries simply because such proteins had not been eluted from the resin, as elegantly demonstrated by Di Girolamo et al. (2011).

References

Adamczyk B, Tharmalingam T, Rudd PM (2012) Glycans as cancer biomarkers. Biochim Biophys Acta 1820:1347–1353

An HJ, Froehlich JW, Lebrilla CB (2009) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. Curr Opin Chem Biol 13:421–426

Bandow JE (2010) Comparison of protein enrichment strategies for proteome analysis of plasma. Proteomics 10:1416–1425

Barnett JP, Scanlan DJ, Blindauer CA (2012) Fractionation and identification of metalloproteins from a marine cyanobacterium. Anal Bioanal Chem 402:3371–3377

Basset C, Dedieu A, Guérin P, Quéméneur E, Meyer D, Vidaud C (2008) Specific capture of uranyl protein targets by metal affinity chromatography. J Chromatogr A 1185:233–240

Bellei E, Bergamini S, Monari E, Fantoni LI, Cuoghi A, Ozben T, Tomasi A (2011) High abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. Amino Acids 40:145–156

Boschetti E, Righetti PG (2008) The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. J Proteomics 71:255–264

Boschetti E, Righetti PG (2012) The quest for biomarkers: are we on the right technical track? Proteomics Clin Appl 6:22–41

Boschetti E, Righetti PG (2013) Low-abundance protein discovery: state of the art and protocols. Elsevier, Waltham

Boschetti E, Candiano G, Righetti PG (2013) Current gel electrophoresis approaches to low-abundance protein marker discovery. In: Issaq HJ, Veenstra TD (eds) Proteomic and metabolomic approaches to biomarker discovery. Academic Press, New York

Butterfield DA, Owen JB (2011) Lectin-affinity chromatography brain glycoproteomics and Alzheimer disease: insights into protein alterations consistent with the pathology and progression of this dementing disorder. Proteomics Clin Appl 5:50–56

Clifton J, Huang F, Gaso-Sokac D, Brilliant K, Hixson D, Josic D (2010) Use of proteomics for validation of the isolation process of clotting factor IX from human plasma. J Proteomics 73: 678–688

D'Amato A, Bachi A, Fasoli E, Boschetti E, Peltre P, Sénéchal H, Citterio A, Righetti PG (2010) In-depth exploration of *Hevea brasiliensis* latex proteome and "hidden allergens" via combinatorial peptide ligand libraries. J Proteomics 73:1368–1380

Darula Z, Sherman J, Medzihradszky KF (2012) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. Mol Cell Proteomics 11:O111.016774

Di Girolamo F, Boschetti E, Chung MC, Guadagni F, Righetti PG (2011) "Proteomineering" or not? The debate on biomarker discovery in sera continues. J Proteomics 74:589–594

Drake RR, Schwegler EE, Malik G, Diaz J, Block T, Mehta A, Semmes OJ (2006) Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. Mol Cell Proteomics 5:1957–1967

Edelmann LJ (2011) Strong cation exchange chromatography in analysis of post-translational modifications: innovations and perspectives. J Biomed Biotechnol 2011:936508

Faca V, Pitteri SJ, Newcomb L, Glukhova V, Phanstiel D, Krasnoselsky A, Zhang Q, Struthers J, Wang H, Eng J, Fitzgibbon M, McIntosh M, Hanash S (2007) Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. J Proteome Res 6:3558–3565

Fonslow BR, Carvalho PC, Academia K, Freeby S, Xu T, Nakorchevsky A, Paulus A, Yates JR (2011) Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. J Proteome Res 10:3690–3700

Gao M, Deng C, Yu W, Zhang Y, Yang P, Zhang X (2008) Large scale depletion of the high-abundance proteins and analysis of middle- and low-abundance proteins in human liver proteome by multidimensional liquid chromatography. Proteomics 8:939–947

Gong Y, Li X, Yang B, Ying W, Li D, Zhang Y, Dai S, Cai Y, Wang J, He F, Qian X (2006) Different immunoaffinity fractionation strategies to characterize the human plasma proteome. J Proteome Res 5:1379–1387

Gronborg M, Kristiansen TZ, Stensballe A, Andersen JS (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. Mol Cell Proteomics 1:517–527

Hnasko R, Ben-Jonathan N (2003) Prolactin regulation by heparin-binding growth factors expressed in mouse pituitary cell lines. Endocrine 20:35–44

Huang L, Fang X (2008) Immunoaffinity fractionation of plasma proteins by chicken IgY antibodies. Methods Mol Biol 425:41–51

Huhn C, Ruhaak LR, Wuhrer M, Deelder AM (2011) Hexapeptide library as a universal tool for sample preparation in protein glycosylation analysis. J Proteomics 75:1515–1528

Keidel EM, Ribitsch D, Lottspeich F (2010) Equalizer technology. Equal rights for disparate beads. Proteomics 10:1–10

Kirkland PA, Humbard MA, Daniels CJ, Maupin-Furlow JA (2008) Shotgun proteomics of the haloarchaeon *Haloferax volcanii*. J Proteome Res 7:5033–5039

Krapfenbauer K, Fountoulakis M (2009) Improved enrichment and proteomic analysis of brain proteins with signaling function by heparin chromatography. Methods Mol Biol 566:165–180

Kullolli M, Hancock WS, Hincapie M (2008) Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. J Sep Sci 31:2733–2739

Lathrop JT, Fijalkowska I, Hammond D (2007) The bead blot: a method for identifying ligand-protein and protein–protein interactions using combinatorial libraries of peptide ligands. Anal Biochem 361:65–76

Levin Y, Schwarz E, Wang L, Leweke FM, Bahn S (2007) Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. J Sep Sci 30:2198–2203

Li S, Dass C (1999) Iron (III)-immobilized metal ion affinity chromatography and mass spectrometry for the purification and characterization of synthetic phosphopeptides. Anal Biochem 270:9–14

Maa YF, Antia FD, el Rassi Z, Horvath C (1998) Mixed-bed ion-exchange columns for protein high-performance liquid chromatography. J Chromatogr 452:331–345

Madian AG, Myracle AD, Diaz-Maldonado N, Rochelle NS, Janle EM, Regnier FE (2011) Differential carbonylation of proteins as a function of in vivo oxidative stress. J Proteome Res 10:3959–3972

Medvedev A, Kopylov A, Buneeva O, Zgoda V, Archakov A (2012) Affinity-based proteomic profiling: problems and achievements. Proteomics 12:621–637

Motoyama A, Xu T, Ruse CI, Wohlschlegel JA, Yates JR (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem 79:3623–3634

Ni RJ, Shen Z, Yang CP, Wu YD, Bi YD, Wang BC (2010) Identification of low abundance polyA-binding proteins in Arabidopsis chloroplast using polyA-affinity column. Mol Biol Rep 37:637–641

Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S (2003) Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. Proteomics 3:422–432

Porath J, Olin B (1983) Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel immobilized iron and nickel ions. Biochemistry 22:1621–1630

Posewitz MC, Tempst P (1999) Immobilized gallium (III) affinity chromatography of phosphopeptides. Anal Chem 71:2883–2892

Rho JH, Roehrl MH, Wang JY (2009) Tissue proteomics reveals differential and compartment-specific expression of the homologs transgelin and transgelin-2 in lung adenocarcinoma and its stroma. J Proteome Res 8:5610–5618

Righetti PG, Boschetti E (2012) Breakfast at Tiffany's? Only with a low-abundance deep proteomic signature! Electrophoresis 33: 2228–2239

Righetti PG, Boschetti E, Zanella A, Fasoli E, Citterio A (2010a) Plucking, pillaging and plundering proteomes with combinatorial peptide ligand libraries. J Chromatogr A 1217:893–900

Righetti PG, Boschetti E, Kravchuk A, Fasoli E (2010b) The proteome buccaneers: how to unearth your treasure chest via combinatorial peptide ligand libraries. Expert Rev Proteomics 7:373–385

Righetti PG, Boschetti E, Fasoli E (2011) Capturing and amplifying impurities from recombinant therapeutic proteins via combinatorial peptide libraries: a proteomic approach. Curr Pharm Biotech 12:1537–1547

Righetti PG, D'Amato A, Fasoli E, Boschetti E (2012) In *tabernaquandosumus*: a drunkard's cakewalk through wine proteomics. Food Technol Biotechnol 50:253–260

Rivers J, Hughes C, McKenna T, Woolerton Y, Vissers JP, Langridge JI, Beynon RJ (2011) Asymmetric proteome equalization of the skeletal muscle proteome using a combinatorial hexapeptide library. PLoS One 6:e28902

Roux-Dalvai F, Gonzalez de Peredo A, Simó C, Guerrier L, Bouyssié D, Zanella A, Citterio A, Burlet-Schiltz O, Boschetti E, Righetti PG, Monsarrat B (2008) Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. Mol Cell Proteomics 7:2254–2269

Scholten A, Poh MK, van Veen TA, van Breukelen B, Vos MA, Heck AJ (2006) Analysis of the cGMP/cAMP interactome using a chemical proteomics approach in mammalian heart tissue validates sphingosine kinase type 1-interacting protein as a genuine and highly abundant AKAP. J Proteome Res 5: 1435–1447

Shahali Y, Sutra JP, Fasoli E, D'Amato A, Righetti PG, Futamura N, Boschetti E, Sénéchal H, Poncet P (2012) Allergomic study of cypress pollen via combinatorial peptide ligand libraries. J Proteomics 77:101–110

Shimada Y, Fukuda T, Aoki K, Yukawa T, Iwamuro S, Ohkawa K, Takada K (2008) A protocol for immunoaffinity separation of the accumulated ubiquitin-protein conjugates solubilized with sodium dodecyl sulfate. Anal Biochem 377:77–82

Stone MD, Chen X, McGowan T, Bandhakavi S, Cheng B, Rhodus NL, Griffin TJ (2011) Large-scale phosphoproteomics analysis of whole saliva reveals a distinct phosphorylation pattern. J Proteome Res 10:1728–1736

Su L, Cao L, Zhou R, Jiang Z, Xiao K, Kong W, Wang H, Deng J, Wen B, Tan F, Zhang Y, Xie L (2013) Identification of novel biomarkers for sepsis prognosis via urinary proteomic analysis using iTRAQ labeling and 2D-LC-MS/MS. PLoS One 8:e54237

Thulasiraman V, Lin S, Gheorghiu L, Lathrop J, Lomas L, Hammond D, Boschetti E (2005) Reduction of concentration difference of proteins from biological liquids using combinatorial ligands. Electrophoresis 26:3561–3571

Tian Y, Zhang H (2010) Glycoproteomics and clinical applications. Proteomics Clin Appl 4:124–132

Westbrook JA, Yan JX, Wait R, Welson SY, Dunn MJ (2001) Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms. Electrophoresis 22:2865–2871

Xu Y, Wang BC, Zhu YX (2007) Identification of proteins expressed at extremely low level in Arabidopsis leaves. Biochem Biophys Res Commun 358:808–812

Yang Z, Harri LE, Palmer-Toy DE, Hancock WS (2006) Multilectin affinity chromatography for characterization of multiple glycoprotein biomarker candidates in serum from breast cancer patients. Clin Chem 52:1897–1905

Zhang H, Zha X, Tan Y, Hornbeck PV (2002) Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. J Biol Chem 277:39379–39387