Journal of Chromatography A, 1217 (2010) 893–900

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Review

Plucking, pillaging and plundering proteomes with combinatorial peptide ligand libraries

Pier Giorgio Righetti[a,*], Egisto Boschetti[b], Alberto Zanella[c], Elisa Fasoli[a], Attilio Citterio[a]

[a] Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Via Mancinelli 7, Milano 20131, Italy
[b] Bio-Rad Laboratories, 3, Boulevard R. Poincaré, 92430 Marnes-la-Coquette, France
[c] Hematology 2, Fondazione IRCCS Ospedale Maggiore, Mangiagalli e Regina Elena, via F. Sforza 28, 20122 Milano, Italy

ARTICLE INFO

Article history:
Available online 1 September 2009

Keywords:
Ligand library
Peptide ligands
Low-abundance proteome
Elution systems

ABSTRACT

Recent developments in the technique of combinatorial peptide ligand libraries, for enhancing the visibility of the low-abundance proteome, are reviewed here. Novel en bloc elution systems, allowing essentially complete proteome recovery in a single step, are reported here, particularly, en bloc elution with 3–5% boiling sodium dodecyl sulphate (SDS) or in urea–thiourea–CHAPS added with either 40 mM formic acid or 25 mM cysteic acid. Novel capturing systems are also discussed: in particular, although capturing at pH 7.2 in physiological saline has always been recommended, it is shown that capturing also at acidic (pH 3.8) and alkaline (pH 9.5) values substantially increments the total captured protein population. Some examples of detection of novel proteins by the described methodology are also discussed. In particular, in the case of venom proteins, where essentially all components had been detected and fully described by conventional means, the application of the ligand library technology allowed the discovery of two, previously unreported, trace enzymes necessary for the maintenance of the native structure of venom components, namely peroxiredoxin and glutaminyl cyclase. In the case of the red blood cell (RBC) cytoplasmic proteome, where a grand total of 1570 components of the 2% minority proteomes have been identified, these findings allowed to unravel the genetic defect of a rare RBC disease, called congenital dyserythropoietic anemia type II. The mutations are located in the SEC23B gene coding for the SEC23B protein, detected for the first time in the RBC proteome thanks to the peptide capturing technology.

Contents

1. Prologue

As odd as it sounds, pillaging, plucking and plundering (PPP) has been the favourite past time of all past "civilizations", to the point that the history of human kind is the account of all disasters provoked by infinite wars and the ensuing PPPs perpetrated by the winners onto the losers (unfortunately it does not seem that today we are doing any better!). Perhaps one of the most chilling PPP events took place in 1527 in Rome, and in fact it went down to history as "il sacco di Roma" (the sack of Rome); it is worth recalling some of the episodes here. In the 16 century Europe was shaken on its foundations by some major events, among which the fierce

* Corresponding author.
E-mail address: piergiorgio.righetti@polimi.it (P.G. Righetti).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.08.070

fighting between the French (François I) and the German (Charles V) rulers for the control of the continent. In the long run, Charles V defeated the French King in Pavia in 1524, from where he was taken prisoner to Madrid. In an attempt at revenge, François I called upon the Sultan Suleiman to invade Hungary, a Province of Charles V empire. The Vatican had sided with the French and their policy so, as soon as Charles V could solve his problems at home, he let his general von Frundsberg and his terrifying army of landsknechts to march upon Rome to punish the ruling pope (Clemente VII). No troops could contrast their march. As the 10,000 ferocious soldiers, who had not received their salary for a whole year, hit Rome, havoc ensued. The population at large was massacred and soon 10,000 corpses paved the streets while another 2000 floated on the Tibur river. San Peter and the Vatican were savagely sacked and all roman palaces were pillaged too. Churches were transformed into army barracks and brothels and even the famous Stanze of Raphael in the Vatican used as stables. The barbarian troops even proclaimed pope Martin Luther and it was only by a miracle that the real pope, Clemente VII, survived, as he had barricaded himself in Castel Sant'Angelo, the only structure that the landsknechts could not conquer. Interestingly, the famous artist Benvenuto Cellini was among the guards defending the castle and he even claimed that he shot to death general von Frundsberg, the chief commander of the landsknechts. Attempts by the French and English armies to liberate the pope failed and Clemente VII himself had to escape from Rome taking refuge in Orvieto and then Viterbo. Basically, the pope was held hostage by the German emperor for two years; by the end of this disastrous pillaging, Rome had been left in total ruins.

2. Introduction

Notwithstanding the pernicious effects of PPPs, this procedure seems to draw proselytes from the four corners of the word in the field of proteomics. All of us in this arena dream of pillaging proteomes and running away with the booty, be it a biomarker for a given pathology, a membrane receptor, a novel key enzyme involved in an important biological pathway, a hidden allergen previously undetected. To that purpose, a host of techniques have been devised for ransacking proteomes and bringing to the limelight those low- to very-low-abundance species that constitute its vast majority [1–3]. One technique that is taking momentum is the combinatorial peptide ligand library that is the object of the present review. The story of a combinatorial peptide library commenced at the end of 1980s, when Furka et al. [4] presented the principle as a way to simplify the preparation of peptides. It was about the same period when combinatorial chemistry became successful as an enormous source of new molecules that could be screened for their potential therapeutic application. Within the same context, peptides and peptoids became also a focus for similar applications. With Lam et al. [5], however, peptides from combinatorial libraries were studied for the first time as a possible source of affinity ligands for the separation of proteins from complex mixtures. While the promise of discovering new effective drug candidates from combinatorial libraries was not as successful as hoped, the purification of peptide ligands was possible in a large number of cases. The use of similar libraries in proteomic investigations for detecting many more proteins and especially those with a concentration below the sensitivity of current analytical methods was described for the first time only in 2005 [6]. The interest of such an approach was seen as a novel way to detect not only hidden proteins from complex proteomes, due to the presence of highly abundant species, but even more importantly because it rendered possible the concentration of low- and very-low-abundance species. The interest of such a possibility is large not only for checking the function of rare species present in a given extract or cell organelle, but also for seeing if the genome is still active for proteins that are supposed to be repressed in their expression after the fetal stage. An additional benefit is to find markers of diagnostic interest as well as markers to follow the efficacy of a therapeutic treatment or to classify populations between respondents and non-respondents or even fatal outcome. The possibility of detecting traces of biomarkers permits to diagnose misregulated expressions at earlier stages than with current methods.

We have now applied the ligand library capturing technique to quite a number of biological samples, such as urine [7], serum [8,9], human platelets [10], red blood cells [11], bile fluid [12], chicken egg white [13] and yolk [14], recombinant DNA products [15,16] and snake venoms [17]. In some instances, the use of peptide libraries has enabled us to detect allergens in food stuff (notably maize and cow's whey) that had not been revealed up to the present [18,19]. In all cases, the increment in number of species detected was from two- to five-fold as compared to a control, non-treated sample. This substantial increment in detection applies mostly to low-abundance proteins, considering that they could not be detected in the untreated samples, where they should have been obviously present. In some cases, such as with the human red blood cell (RBC) cytoplasmic proteome, the increment was spectacular: a total of 1570 proteins could be identified [11], as opposed to 250 in the best and deepest proteomic analysis previously published [20]. We have also reviewed the field in quite a large number of articles, to which the reader is referred for a better insight on the technique [21–29]. In one instance, we have described a detailed protocol for solid-phase peptide library users [30]. In another couple of investigations, we have explored the capturing ability of these libraries as a function of the oligopeptide length [31] and even when using just single amino acids attached to the beads [32]. Therefore, in the present review, we will cover only the most recent advances not yet described in all of our previous reviews. Data presented are the result of the use of a commercially available library called ProteoMiner (primary amine terminal hexapeptides from Bio-Rad Laboratories, Hercules, CA, USA) and from a carboxylated library specifically chemically modified and that is not commercially available as we write.

3. General properties and behaviour of hexapeptide ligand libraries

As we have extensively described the properties and behaviour of hexapeptide libraries, we will here only briefly summarize their general characteristics. The solid-phase product is a mixture of porous beads on which hexapeptides are covalently attached. Each bead carries a large number (billions) of copies of the same peptide bait; the beads are thus different from each other, and all combinations of hexapeptides are present. Depending on the number of amino acids used, a library contains a population of millions of different ligands (*e.g.* 11, 24 or 64 millions starting, respectively, from 15, 17 or 20 different amino acids). When a complex protein extract is exposed to such a ligand library in large overloading conditions, each bead with affinity to an abundant protein will rapidly become saturated, and the vast majority of the same protein will remain unbound. In contrast, trace proteins will not saturate the corresponding partner beads, but are captured in progressively increasing amounts as the beads are loaded with additional protein extract. Thus, a solid-phase ligand library enriches for trace proteins, while concomitantly reducing the relative concentration of abundant species. In theory each bead that carries a single peptide ligand should interact with proteins that share the same epitope complementary to the peptide bait. However, because in a number of cases peptide ligands can differ from each other by just one amino acid, similar interactions can be found with more than one protein—the single interactions are governed in this case by differ-

ent dissociation constants. As experimentally demonstrated [23], in fact, the content of one single bead (single peptide structure) after contact with a crude protein extract, is constituted of few proteins—some dominant and other minor species. Generally, the affinity between a hexapeptide ligand for a given protein might be considered as a rather weak binding event; however, experience with millions of hexapeptides has demonstrated that indeed such complexes can be very stable, sometimes requiring harsh conditions for the dissociation, such as 6 M guanidine hydrochloride.

4. Novel elution protocols

Hexapeptide ligand libraries as a mean of capturing proteins represent a sort of synthesis of all possible modes of interactions [21] that are commonly observed when used separately. Due to this complex interaction situation the methods of elution are relatively unusual when trying to recover the largest number of proteins. The goal is in fact to elute absolutely every single protein species captured even if very tightly anchored to the affinity bait as a consequence of an extremely large affinity constant. To this end proteins are sequentially eluted with rather strong desorbing agents, able to disrupt the various bonds implicated in the ligand–protein binding events (typically hydrogen, ionic, hydrophobic and dipolar interactions that occur in a fully native configuration, in physiological saline) [33]. Therefore, when an absolute elution process is desired, sequential steps are used. The two most popular eluting agents are, in the order, TUC (7 M urea, 2 M thiourea and 3% CHAPS) followed by UCA (9 M urea in 50 mM citric acid, pH 3.3, 2% CHAPS). Quite often, a third eluant is used (hydro-organic eluting mixture, *i.e.* 6% (v/v) acetonitrile, 12% (v/v) isopropanol, 2% (v/v) of ammonia and 80% (v/v) water). In a few cases, our group has adopted even a sequence of four eluants, the three ones reported above, preceded by 1 M NaCl. In some circumstances it is also recommended to use other eluting agents such as 40–60% ethylene glycol for strong hydrophobic interaction and a deforming agent such as 200 mM glycine–HCl, pH 2.4, as adopted in most immuno-affinity chromatography situations. Table 1 summarizes various elution methods and related references for additional details.

Elution agents could also be selected as a function of the following protein analysis. For instance if the purpose is to digest the entire set of captured species, instead of eluting, a trypsin treatment of proteins can directly be operated on the beads. Conversely when a SDS-PAGE analysis will have to be performed in view of gel slicing for an in depth LC–MS/MS analysis, protein elution from the beads can be performed directly by 5–10% SDS (see below). Although these desorbing solutions allow discovering an incredible number of proteins undetectable by any other technique, clearly the use of a sequence of three or four elution steps entails a considerable amount of extra workload. In an attempt at reducing the complexity of these desorption methods, we have recently reported [34] a single and very convenient, en bloc elution, able to release at once, almost quantitatively, all proteins captured by the beads: sodium dodecyl sulphate (SDS), under reducing and boiling conditions, as customarily performed for SDS-PAGE [35]. Eluted proteins could easily be used for current analysis such as two-dimensional electrophoresis and LC–MS/MS. Fig. 1 shows representative 2D maps of native normal dialyzed human urine proteins (panel A, 400 mL urine) and the same sample (dialyzed "B" or not dialyzed "C") after treatment with peptide beads where it is demonstrated that a preliminary dialysis is very useful to evidence a larger number of spots. The reason for the improved data is related to the elimination of interfering small molecules as well as the normalization of ionic strength and pH to PBS. The last two maps were taken from peptide library beads eluted in boiling 10% SDS added with 3% DTE. After elimination of excess SDS in chloroform/methanol, the pellet was re-suspended in reduction-alkylation mixtures and run in a 2D map by adopting as a first dimension a non-linear, pH 3–10 IPG "soft" (3.3%T, 3.5%C) strip. As shown in panel A (left side), the protein pattern includes high-abundance proteins such as albumin, heavy and light chains of immunoglobulins as well as other proteins derived from serum, such as α_1-microglobulin. The anal-

Table 1
Protein elution methods as a function of application or type of analysis:.

Type of sample analysis	Elution method	Comments	References
SDS-PAGE	Single elution with 10% sodium dodecyl sulphate containing 50 mM DTT	Followed or not by LC–MS/MS (gel slices followed by trypsination) or Western blot	[34]
2D-PAGE	2 M thiourea–7 M urea–4% CHAPS–50 mM cysteic acid 9 M urea–4%CHAPS–100 mM acetic acid, pH 3 9 M urea–4%CHAPS–100 mM ammonia, pH 11 6 M guanidine–HCl, pH 6.0 Various sequences (see references)	Thiourea–urea–CHAPS–cysteic acid fully compatible with 2D-PAGE analytical method Eluates from urea–CHAPS and guanidine –HCl require a clean-up step to reduce the presence of salts	[30]
SELDI-TOF	9 M urea containing 2% CHAPS at pH 3.0–3.5 with 50 mM citric acid (or other strong acids) 6M guanidine–HCl pH 6 2 M thiourea–7 M urea–4% CHAPS–50 mM cysteic acid 9 M urea–4%CHAPS–100 mM ammonia, pH 11	Described methods are fully compatible with this application	[6,7]
IEF	See 2D-PAGE above	Eluates from all methods require a clean-up step to reduce the presence of salts	[19,30]
2D-DIGE	20 mM Tris containing 7 M urea, 2 M thiourea and 4% CHAPS pH 8.5 (sodium carbonate could also be used instead of Tris)	The elution buffer is compatible with cyanine dye labelling protocols	[44]
ELISA-based assays	0.2 M glycine–HCl, 2% NP-40, pH 2.4; 0.1 M acetic acid, 2% NP-40; 1 M NaCl, 2% NP-40; or 0.1 M acetic acid containing 40% ethylene glycol	Harsh elution conditions may denature captured proteins rendering them unable to react with antibodies. Each indicated elution method may not desorb all captured proteins. They could be used sequentially and eluates pooled and neutralized with 3 M Tris base	Unpublished data
Direct trypsin digestion for LC–MS/MS	No elution required. Direct trypsin treatment of beads loaded with captured proteins yields a supernatant directly used for sequencing	After direct trypsin digestion beads are separated by centrifugation and supernatant concentrated and added with formic acid	[16]

Abbreviations: SDS: sodium dodecyl sulphate; PAGE: polyacrylamide gel electrophoresis; DIGE: difference in-gel electrophoresis; DTT: di-thiothreitol; LC–MS/MS: liquid chromatography–tandem mass spectrometry; SELDI-TOF: surface enhanced laser desorption ionization–time of flight; IEF: isoelectric focusing; ELISA: enzyme linked immunosorbent assay.

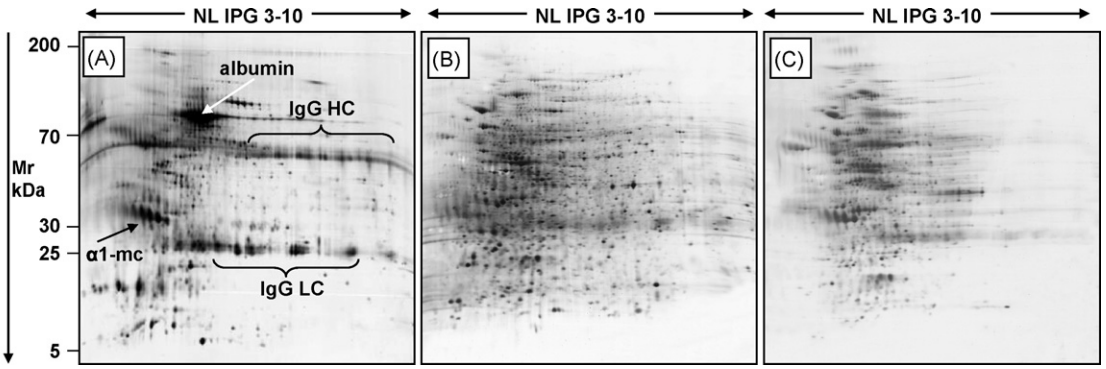


Fig. 1. 2DE pattern comparison of dialyzed total urinary proteins without any treatment (A) or with the same sample treated with hexapeptide ligand library dialyzed (B) and not dialyzed (C). The Last two samples were eluted with a single step in SDS-DTE. The major protein components – albumin, IgG light (IgG LC) and heavy chains (IgG HC) and α_1-microglobulin (α_1-mc) are indicated. The pH gradient was between 3 and 10, non-linear. The protein detection was with silver stain (from Candiano et al. [34], by permission).

ysis of treated sample evidences a very different pattern with the reduction of high-abundance species and the appearance of many other low-abundance species (see panel B) undetectable in native, untreated urine. Overall an incredible array of well focused, protein spots (>1000) could be displayed in panel B, covering the entire pH interval and ranging from 10 to >200 kDa in size.

Although not yet experienced SDS elution could be of interest to the recovery of hydrophobic proteins and proteins from leaf extracts.

In another approach, in order to avoid the post-treatment of SDS elution, namely the removal of the surfactant from the protein solution, we have reported two additional en bloc elution systems, which are able to achieve recoveries >96% in a single step [36]. One consists of TUC (7 M urea, 2 M thiourea, 3% CHAPS) added with 40 mM formic acid, the other of TUC added with 25 mM cysteic acid (Cys-A). Although both systems are almost equally performing, the formic acid eluant has as a drawback, namely the potential to modify proteins by formylation of Ser and Thr residues. On the contrary, the Cys-A system is unreactive towards proteins. Additionally, Cys-A, due to its very low p*I* value (1.80) does not interfere with subsequent 2D map analyses since, during the first isoelectric focusing step (in general performed in immobilized pH gradients), it migrates to the anodic compartment and thus vacates the gel. Conversely, formic acid would mostly collect around pH 3 and acetic or citric acid, formerly used in the UCA (9 M urea, 50 mM citric acid) eluant, would condense around pH 4 in the focusing step, interfering thus with 2D map analyses. We believe that, with these three novel elution systems reported, a much greater versatility on the use of the described technology can be achieved.

5. Proteome capture at different pH values

Although combinatorial hexapeptide ligand libraries are a highly performing tool for detecting the low-abundance proteome, a disturbing feature we have noticed upon their continuous use is the fact that they are not “ecumenical” as we thought they would be, since, although we can detect plenty of invisible species, we also lose some of the components of the initial mixture, a loss that can reach 10% and even a bit higher. This was evidenced for several analysed proteomes [10,11,13] by comparing the lists of gene products found by LC–MS/MS before and after treatment with ligand libraries. Some gene products that were found in the initial sample were not detectable after treatment. Up to the present, a partial remedy we found to this defect was to use two types of libraries, the standard one (ProteoMiner) in which all hexapeptides terminate with a free amino group, and a second library in which all termini are carboxylated. With this last library we seem to be able to capture an additional, diverse population that reduces the losses in the initial sample [13,14]. We thought there could be an additional way to reduce the initial sample loss, namely to perform the capture at different pH values. Let us consider the following: the interaction between proteins and hexapeptide library is a full thermodynamic phenomenon. All proteins interact with a single hexapeptide with affinity constants ranging from very low values (virtually zero) to extremely large numbers that could by higher by several orders of magnitude. At the same time one protein likely might interact with many different types of hexapeptides with also affinity constant ranging from very low values up to extremely large numbers. In addition to this intricate situation, when dealing with very complex mixtures such as human serum, there is a strong competition between proteins and hexapeptides with displacement effects that depend on the concentration of given proteins. In other words diluted proteins with high affinity constants will also displace proteins that are concentrated but with a relatively weak affinity for the same peptide. Since the mass action law depends largely also on the pH (the concentration of protons changes the configuration of proteins and modifies the affinity constant for hexapeptides), as a consequence there could be proteins that increase their affinity for some hexapeptide structures and others that lose the capability to interact with the same peptides. Therefore, by using different pH values the entire protein pattern could change significantly at least for a certain number of proteins. These changes may suggest specific practical applications: (i) the proteome could be more completely investigated when exploring the pH scale rather than using a single pH (e.g. neutral conditions) where we know that about 10–15% of species are lost; (ii) when focusing on a single protein species it could be possible to select peptides that capture the proteins more specifically and release it when modifying the environmental pH; (iii) it could be possible to make a distinction between direct and indirect (protein–protein) interaction among proteins and the peptides of the library.

All experiments performed so far, have been operated at pH 7.2, equilibrated in physiological saline. At the light of the above considerations, the capture of the initial sample at additional pH values was performed, namely at pH 4.0 and at pH 9.3. Internal environmental pH of each column was maintained by the presence of solid-state buffers so that the capture was performed just using low concentration of sodium chloride with three columns in series. The results with human sera are shown in Fig. 2A: it can be appreciated that compared to the initial non-treated serum (map A) whatever the pH (pH 7.2 “B”; pH 4.0 “C”; pH 9.3 “D”) more spots were detectable as expected. In addition it can be observed that some exclusive spots can be detected at acidic pH (red circles) and also at alkaline pH (blue circles). Both sets of exclusive spots add to the spots obtained when the capturing pH is around 7.2 and therefore allow increasing the number of discovered species

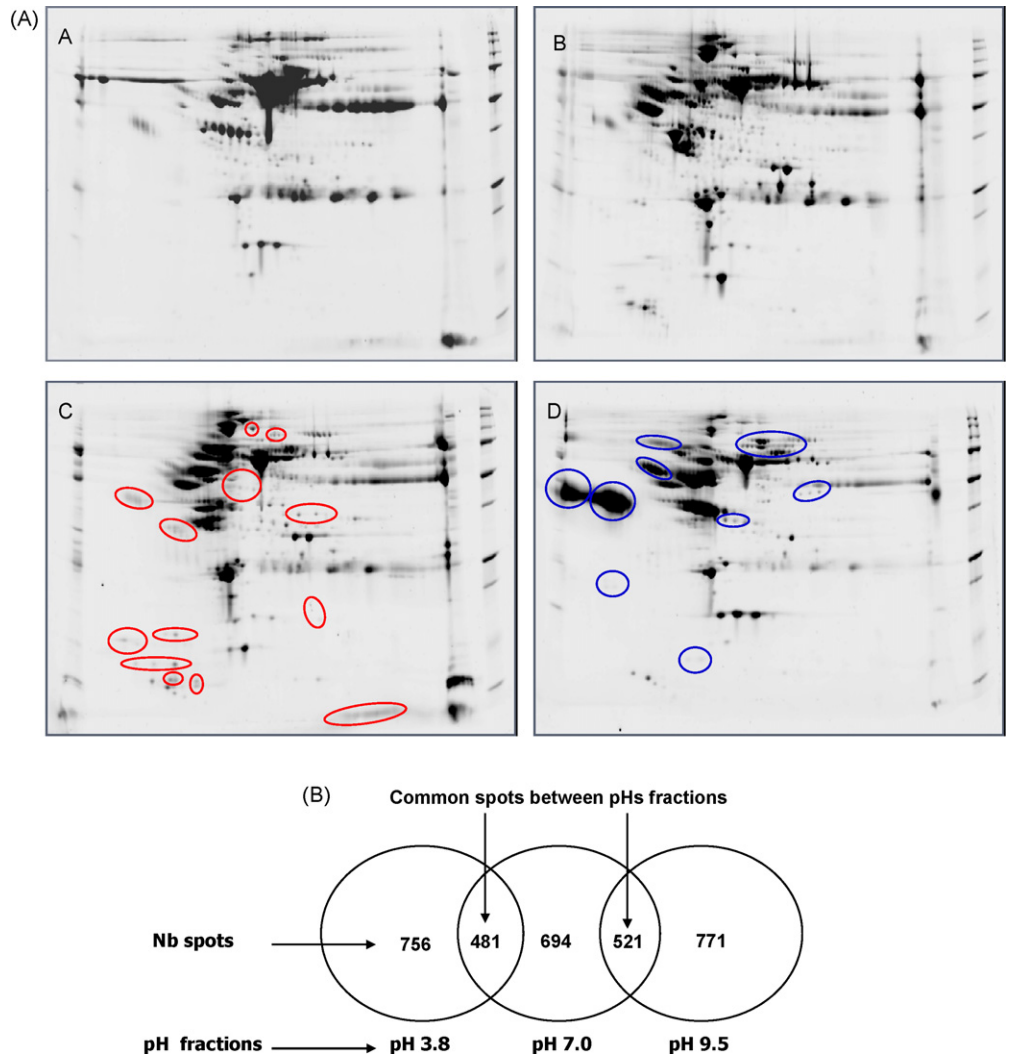


Fig. 2. (A) Two-dimensional maps of initial human sera (panel A) and after capture at pH 7.2 (panel B), pH 4.0 (panel C) and pH 9.3 (panel D). Coomassie blue staining (by courtesy of Dr Cheng Jun Sun, Bio-Rad Laboratories). Coloured circles of lower panels indicate exclusive proteins found at extreme pHs. (B) Overlapping Venn diagrams of differential spot counts performed by PDQuest analysis of silver-stained gels.

(unpublished data). In another experiment similar gels have been overstained with silver and subjected to total spot count and differential analysis with the PDQuest software: as shown in Fig. 2B, a substantial number of additional spots can be counted, drastically reducing the initial sample loss. Thus, these triple capture conditions at these 3 pH values could become a standard protocol in all future peptide library applications.

6. Examples of some biological applications

As stated in Section 2, we will report here only some novel biological applications that have produced outstanding results and have not been treated in previous reviews. For detailed examples of previous work, the readers are referred to the original articles on the various biological samples [7–19] or to previous reviews highlighting some of these findings.

6.1. Snake venoms

We have recently started a collaboration with Dr. Calvete's lab in Valencia (Spain), a world renowned lab for outstanding research in snake venoms [37,38]. When we proposed to analyse some venoms to a deeper extent via the described technology, we were met with scepticism, in that we were told that all the possible proteinaceous components in snake venoms had been already fully detected and mapped. Nevertheless, we started a project on the proteomic characterization of the venom of the North American western diamondback rattlesnake, *Crotalus atrox*, by using two complementary approaches: snake venomics (to gain an insight of the overall venom proteome), and two solid-phase combinatorial peptide ligand libraries (regular primary amine terminal hexapeptide and carboxylated form of the same), followed by 2D electrophoresis and mass spectrometric characterization of in-gel digested protein bands (to capture and "amplify" low-abundance proteins). The venomics approach revealed ~24 distinct proteins belonging to two major protein families (snake venom metalloproteinases (SVMP) and serine proteinases) representing 69.5% of the total venom proteins, four medium abundance families (medium-size disintegrin, PLA_2, cysteine-rich secretory protein, and L-amino acid oxidase) amounting to 25.8% of the venom proteins, and three minor protein families (vasoactive peptides, endogenous inhibitor of SVMP, and C-type lectin). This toxin profile potentially explains the cytotoxic, myotoxic, hemotoxic, and hemorrhagic effects evoked by *C. atrox* envenomation. On the other hand, the two combinatorial hexapeptide libraries captured distinct sets of proteins. Beyond a large number of common proteins, carboxylated hexapeptides library contributed to extend the capture to

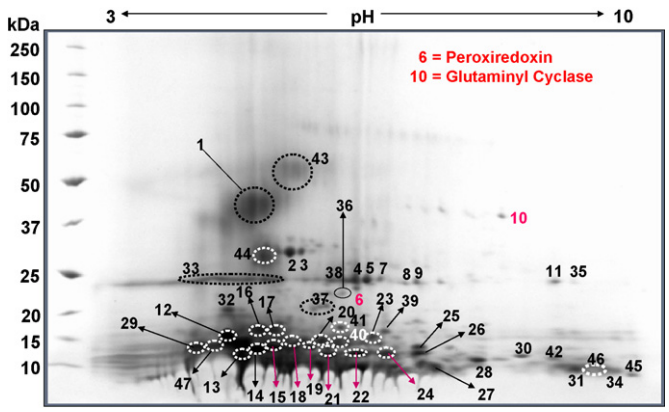


Fig. 3. Colloidal Coomassie blue-stained 2DE map of *C. atrox* venom eluate from the N-termini ligand library. Two hundred micrograms of total proteins were loaded in the first dimension strip. All the circled and numbered spots have been eluted and identified by mass spectrometry. Notice that spots nos. 6 and 10 represent two low-abundance enzymes (respectively peroxiredoxin and gutaminyl cyclase) never previously detected in snake venoms (modified from Calvete et al. [17]).

additional alkaline species. In the library-treated eluates a number of protein species barely, or not at all, detectable in the whole venom were enriched. The amplified low copy number *C. atrox* venom proteins comprised a C-type lectin-like protein, several PLA_2 molecules, PIII-SVMP isoforms (found exclusively from carboxylated library eluate), a 2-cys peroxiredoxin (found exclusively from regular library eluate) and, glutaminyl cyclase (found from both library eluates). Peroxiredoxin and glutaminyl cyclase may participate, respectively, in redox processes leading to the structural/functional diversification of toxins, and in the N-terminal pyrrolidone carboxylic acid formation required in the maturation of bioactive peptides such as bradykinin-potentiating peptides and endogenous inhibitors of metalloproteases. These two enzymes, never previously detected in snake venoms, have been found for the first time thanks to the peptide library capturing technology (see Fig. 3, spots nos. 6 and 10). These findings underscore the usefulness of combinatorial peptide libraries as powerful tools for mining below the tip of the iceberg**,** complementing thereby the data gained using the snake venomics protocol towards a complete visualization of the venom proteome [17].

6.2. Newer insights on the red blood cell cytoplasmic proteome

We have already dealt with this topic on the RBC cytoplasmic proteome in the original article [11] as well as in a recent review [27]. Yet some outstanding, novel and unexpected results call for revisiting this topic. It was an absolute surprise for us to find a grand total of 1570 proteins in the minority proteome, while in the most extensive exploration published by Mann's group [20] only 250 species could be described. Just to give an impression on these data, we re-propose here Fig. 4, in which the unique catch in the right panel might appear to have been painted over with a fine brush by the famous Spanish artist Juan Mirò (yet it is real). How to deal with this large body of proteins, most of them detected for the first time in RBCs was not readily apparent. We enucleated all the globin chains identified and made some general considerations on them. After ProteoMiner treatment, no fewer than eight different globin chains could be detected. The first two are well-known components of adult Hb, which is a tetramer composed of two α- and two β-chains. The two following ones (γ- and ε-globins) are known as fetal chains, whose genes were believed to be silenced a few months after birth. Our data demonstrated that, in reality, they continue to produce globin chains, at a level well below 1% as compared to the α- and β-chains, all along the entire life span. Nobody, however, had ever reported or even suspected the presence of another four globin chains (ζ, θ, δ, μ, called embryonic chains) in an adult, because then genes were thought to be silenced already in the switch from embryo to a developed foetus. These findings pose the intriguing question: if genes are truly silenced or are kept in a "state of alert" so as to be ready for any emergency, should the need arise. We thus, just like the famous "jumping genes" of McClintock [39], have nicknamed them "Stakhanovite genes" that, like the Russian working hero, never quite stop working. It would be nice to see if this hypothesis would be verified by ProteoMiner treatment of other cell lysates, in search of the expression of "silent genes".

But there is more to it. Geneticists and molecular biologists started delving into this long list of proteins and came out with some remarkable results. It turns out that there exist a rather rare RBC defect, called congenital dyserythropoietic anemia (CDA), which in reality encompasses a group of rare hereditary disorders characterized by ineffective erythropoiesis and distinct morphological abnormalities of erythroblasts in the bone marrow. Three major types (I, II and III) and several minor subgroups have been identified [40,41]. CDA II, transmitted as an autosomal recessive

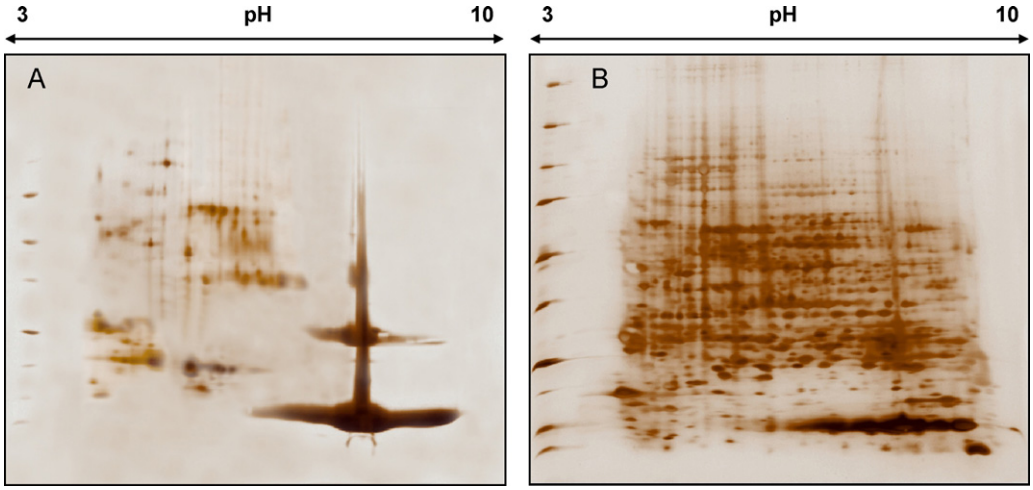
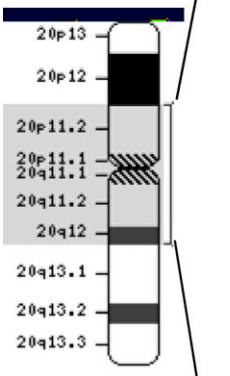


Fig. 4. 2D analysis of proteins from a lysate of pure human red blood cells. The initial extract "A" is represented on the left and shows about 80 spots on a load of 1300 μg of total proteins. The mixture of all eluates from the ligand library column is shown on the right "B" with a number of counted spots of >1000 on a load of 640 μg of total proteins (two times less than control). First dimension: linear IPG pH 3–10; second dimension: 8–18%T polyacrylamide gel. Mr = mass markers. Gels are stained with silver (unpublished experiments with C. Simò).

trait, is the most frequent with more than 300 patients reported. The clinical picture is characterized by mild to moderate anemia, jaundice, variable splenomegaly and iron overload as a consequence of increased red cell turnover due to both ineffective erythropoiesis and peripheral hemolysis. In CDA II, 10–40% of erythroblasts are bi- or multinucleated. A consistent finding in CDA II is a relevant hypoglycosylation of erythrocyte membrane proteins: in particular, band 3 (anion exchange protein 1), the most abundant protein, displays a sharper and faster migrating band on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). It is a matter of debate if the impaired glycosylation is the primary cause of the disorder or a phenomenon secondary to other pathogenic mechanisms. Recent studies on N-glycosylation of erythrocyte membrane proteins in CDA II patients indicate that CDA II is not a distinct glycosylation disorder but is caused by a defect disturbing Golgi processing in erythroblasts; however, the biochemical mechanism causing CDA II remains unknown and the aberrant gene has not been so far elucidated. A linkage analysis located a candidate region spanning 5 cM on chromosome 20 (q11.2), termed CDAN2 locus, in the majority of CDA II patients; however, several candidate genes have been excluded so that up to the present the true genetic defect has not been discovered. Among the 1578 proteins previously found to be expressed on the cytoplasmic proteome of red blood cells [11], Bianchi et al. [42] identified 17 proteins codified by genes located in the chromosomic region between 20p11.23 and 20q11.23: *SNX5*, *SEC23B*, *DTD1*, *NAT5*, *GINS1*, *BCL2-L1*, *MAPRE1*, *CHMP4B*, *EIF2S2*, *AHCY*, *ACSS2*, *GSS*, *EIF6*, *CPNE1*, *EPB41L1*, *C20orf77*, *TGM2* (see Fig. 5). After a number of considerations and known pathological expressions of CDAII, 16 of them were excluded and the attention focused on *SEC23B*, based on its possible role in the endoplasmic reticulum-to Golgi trafficking (OMIM GeneID: 610512) and its localisation on 20p11, the region with the highest LOD-score in CDA II after the recent mapping of the markers on the current contig [43]. The 20 exons and intronic flanking regions of *SEC23B* gene were analysed by direct sequencing in 13 CDA II patients from 10 families; 12 different mutations were detected among the 25 mutated alleles identified: six of them were missense, 2 frameshift, 1 splicing and 3 stop codon. All the missense mutations affected highly conserved amino acids, and were not found in 100 normal alleles examined. Two of them (c.40C > T and c.325G > A) were detected in various unrelated patients. The *SEC23B* protein is a member of the SEC23/SEC24 family, a component of COPII coat protein complex which is involved in protein trafficking through membrane vesicles. Even if the exact function of human *SEC23B* is not completely clarified, abnormalities in this gene are likely to disturb ER-to Golgi trafficking affecting different glycosylation pathways and ultimately accounting for the cellular phenotype observed in CDAII. This is a unique finding in that, for the first time, a genetic defect could be elucidated via proteomic tools, rather than via genetic means, as commonly performed up to the present.

7. Discussion

We hope we have reinforced in this brief survey the notion that combinations of peptides singularly attached to beads that



Start -stop	symbol	O	Cyto	Description
17870244 -17897490	SNX5	-	20p11	sorting nexin 5
18436188 -18490050	SEC23B	+	20p11.23	Sec23 homolog B (S. cerevisiae)
18516556 -18692561	DTD1	+	20p11.23	D-tyrosyl-tRNA deacylase 1 homolog (S. cerevisiae)
19945937-19962269	NAT5	+	20p11.23	N-acetyltransferase 5
25336323 -25377191	GINS1	+	20p11.21	GINS complex subunit 1 (Psf1 homolog)
29715922 -29774317	BCL2L1	-	20q11.21	BCL2-like 1
30871435 -30901867	MAPRE1	+	20q11.1-q11.23	microtubule-associated protein, RP/EB family, member 1
31862780 -31905831	CHMP4B	+	20q11.22	chromatin modifying protein 4B
32139776 -32163746	EIF2S2	-	20pter-q12	eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa
32331737 -32354784	AHCY	-	20cen-q13.1	S-adenosylhomocysteine hydrolase
32926406 -32979423	ACSS2	+	20q11.22	acyl-CoA synthetase short-chain family member 2
32979897 -33007262	GSS	-	20q11.2	glutathione synthetase
33330139-33336008	EIF6	-	20q12	eukaryotic translation initiation factor 6
33677380 -33716262	CPNE1	-	20q11.22	copine I
34163762 -34284135	EPB41L1	+	20q11.2-q12	erythrocyte membrane protein band 4.1-like 1
36095362 -36154180	C20orf77	+	20q11.21-q12	chromosome 20 open reading frame 77
36190277 -36227114	TGM2	-	20q12	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)

Region Displayed: 17,800K-37,100K bp
Total Genes On Chromosome: 824
Total Genes in Region: 281

Fig. 5. Left side: chromosome 20, region displayed: 17,800–37,100 kbp. This region would be compatible with the 17 gene products listed on the right side, as found in the red blood cell cytoplasmic proteome after ligand library treatment. After a number of considerations, protein Sec23 homolog B (second from top, last column) was selected as the potentially mutated protein in the CDAII pathology (from Bianchi et al. [42], by permission).

are mixed together have much to offer to the proteomics community and that, now that the beads are commercially available as ProteoMiner, scientists should take advantage of this opportunity. There is one point, though, that we would like to highlight: in order to get access to the very-low-abundance proteome, it is imperative to use the mixed beads under large overloading conditions. In other words, one should apply as much sample as possible to the ligand library, so as to harvest as many copies of each gene products as available in the sample under investigation. E.g., if in 1 mL of, say, human serum, there are only 10 copies of a rare protein that could be an excellent marker for disease, even if all of them are harvested by the hexapeptide ligand library, this tiny amount will still be much too low and surely below the detection limits of even the most powerful mass spectrometer. In order to be able to see such a rare protein, one might have to use 10 or 50 mL of this sample, so as to collect enough copies of this rare specimen to allow its detection. A note of caution is however necessary because the "amplification" effect applies to all proteins contained in the sample including contaminants. Thus, possible proteins from other sources (e.g. contaminating biologicals) could appear from the peptide treatment creating confusion. Alternatively the volume of beads could be reduced to 100 μL saving thus 90% of the initial sample; however, one has to consider that the eluate from beads would also account for smaller amount of captured proteins. As an example, it is true that for the first time we could detect peroxiredoxin and glutaminyl cyclase in venoms, but to see such low-level species we had to start with a total sample load of 2 g! We doubt we could have seen them in, e.g., just 10 or 50 mg of total protein load. In the RBC lysate, it is true that we could detect an impressive number of species, as many as 1570 at the latest count, but here too the total initial sample load onto the bead library bed was 6 g (although one should consider that haemoglobin alone would represent 5.88 g of the total)! Therefore overloading (an anathema in most chromatographic approaches) should be the rule, not the exception.

Acknowledgements

PGR is supported by grants from ERG senior investigator grant, by PRIN 2009 (MIUR, Rome) and by Fondazione Cariplo. The authors would like to thank providers of experimental data (Drs. P. Bianchi, C. Simò, C.J. Sun) for their help.

References

[1] P.G. Righetti, A. Castagna, P. Antonioli, E. Boschetti, Electrophoresis 26 (2005) 297–319.
[2] P.G. Righetti, Electrophoresis 25 (2004) 2111–2127.
[3] D. Penque, Proteomics Clin. Appl. 3 (2009) 155–172.
[4] A. Furka, F. Sebestyén, M. Asgedom, G. Dibó, Proceedings of 10th International Symposium on Medical Chemistry, Budapest, Hungary, August 15–19, 1988, p. 288.
[5] K.S. Lam, S.E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski, R.J. Knapp, Nature 354 (1991) 82–84.
[6] V. Thulasiraman, S. Lin, L. Gheorghiu, J. Lathrop, L. Lomas, D. Hammond, E. Boschetti, Electrophoresis 26 (2005) 3561–3571.
[7] A. Castagna, D. Cecconi, L. Sennels, J. Rappsilber, L. Guerrier, F. Fortis, E. Boschetti, L. Lomas, P.G. Righetti, J. Proteome Res. 4 (2005) 1917–1930.
[8] L. Sennels, M. Salek, L. Lomas, E. Boschetti, P.G. Righetti, J. Rappsilber, J. Proteome Res. 6 (2007) 4055–4062.
[9] L. Guerrier, V. Thulasiraman, A. Castagna, F. Fortis, S. Lin, L. Lomas, P.G. Righetti, E. Boschetti, J. Chromatogr. B 833 (2006) 33–40.
[10] L. Guerrier, S. Claverol, F. Fortis, S. Rinalducci, A.M. Timperio, P. Antonioli, M. Jandrot-Perrus, E. Boschetti, P.G. Righetti, J. Proteome Res. 6 (2007) 4290–4303.
[11] F. Roux-Dalvai, A. Gonzalez de Peredo, C. Simò, L. Guerrier, D. Bouyssie, A. Zanella, A. Citterio, O. Burlet-Schiltz, E. Boschetti, P.G. Righetti, B. Monsarrat, Mol. Cell. Proteomics 7 (2008) 2254–2269.
[12] L. Guerrier, S. Claverol, L. Finzi, F. Paye, F. Fortis, E. Boschetti, C. Housset, J. Chromatogr. A 1176 (2007) 192–205.
[13] C. D'Ambrosio, S. Arena, A. Scaloni, F. Fortis, E. Boschetti, M.E. Mendieta, A. Citterio, P.G. Righetti, J. Proteome Res. 7 (2008) 3461–3474.
[14] A. Farinazzo, U. Restuccia, A. Bachi, L. Guerrier, F. Fortis, E. Boschetti, E. Fasoli, A. Citterio, P.G. Righetti, J. Chromatogr. A 1216 (2009) 1241–1252.
[15] F. Fortis, L. Guerrier, P.G. Righetti, P. Antonioli, E. Boschetti, Electrophoresis 27 (2006) 3018–3027.
[16] F. Fortis, L. Guerrier, L.B. Areces, P. Antonioli, T. Hayes, K. Carrick, D. Hammond, E. Boschetti, P.G. Righetti, J. Proteome Res. 5 (2006) 2577–2585.
[17] J.J. Calvete, E. Fasoli, L. Sanz, E. Boschetti, P.G. Righetti, J. Proteome Res. 8 (2009) 3055–3067.
[18] E. Fasoli, E.A. Pastorello, L. Farioli, J. Scibilia, G. Aldini, M. Carini, A. Marocco, E. Boschetti, P.G. Righetti, J. Proteomics 72 (2009) 501–510.
[19] A. D'Amato, A. Bachi, E. Fasoli, E. Boschetti, G. Peltre, H. Sénéchal, P.G. Righetti, J. Proteome Res. 8 (2009) 3925–3936.
[20] E.M. Pasini, M. Kirkegaard, P. Mortensen, H.U. Lutz, A.W. Thomas, M. Mann, Blood 108 (2006) 791–801.
[21] P.G. Righetti, E. Boschetti, L. Lomas, A. Citterio, Proteomics 6 (2006) 3980–3992.
[22] P.G. Righetti, E. Boschetti, FEBS J. 274 (2007) 897–905.
[23] E. Boschetti, L. Lomas, A. Citterio, P.G. Righetti, J. Chromatogr. A 1153 (2007) 277–290.
[24] E. Boschetti, B. Monsarrat, P.G. Righetti, Curr. Proteomics 4 (2007) 198–208.
[25] E. Boschetti, P.G. Righetti, BioTechniques 44 (2008) 663–665.
[26] E. Boschetti, P.G. Righetti, J. Proteomics 71 (2008) 255–264.
[27] P.G. Righetti, E. Boschetti, Mass Spectrom. Rev. 27 (2008) 596–608.
[28] E. Boschetti, L.V. Bindschedler, C. Tang, E. Fasoli, P.G. Righetti, J. Chromatogr. A 1216 (2009) 1215–1222.
[29] E. Boschetti, P.G. Righetti, Proteomics 9 (2009) 1492–1510.
[30] L. Guerrier, P.G. Righetti, E. Boschetti, Nat. Protoc. 3 (2008) 883–890.
[31] C. Simó, A. Bachi, A. Cattaneo, L. Guerrier, F. Fortis, E. Boschetti, A. Podtelejnikov, P.G. Righetti, Anal. Chem. 80 (2008) 3547–3556.
[32] A. Bachi, C. Simó, U. Restuccia, L. Guerrier, F. Fortis, E. Boschetti, M. Masseroli, P.G. Righetti, Anal. Chem. 80 (2008) 3557–3565.
[33] K.E. van Holde, W.C. Johnson, P.S. Ho, Principles of Physical Biochemistry, Prentice-Hall International, London, 1998, pp. 2–15.
[34] G. Candiano, V. Dimuccio, M. Bruschi, L. Santucci, R. Gusmano, E. Boschetti, P.G. Righetti, G.M. Ghiggeri, Electrophoresis 30 (2009) 2405–2411.
[35] A.T. Andrews, Electrophoresis: Theory, Techniques and Biochemical and Clinical Applications, Clarendon Press, Oxford, 1986, pp. 117–144.
[36] A. Farinazzo, E. Fasoli, A.V. Kravchuk, G. Candiano, G. Aldini, L. Regazzoni, P.G. Righetti, J. Proteomics 72 (2009) 725–730.
[37] J.J. Calvete, J. Proteomics 72 (2009) 121–126.
[38] J.J. Calvete, P. Juárez, L. Sanz, J. Mass Spectrom. 42 (2007) 1405–1414.
[39] B. McClintock, Proc. Natl. Acad. Sci. U.S.A. 36 (1950) 344–355.
[40] S.N. Wickramasinghe, W.G. Wood, Br. J. Haematol. 131 (2005) 431–446.
[41] H. Heimpel, Ann. Hematol. 83 (2004) 613–621.
[42] P. Bianchi, E. Fermo, C. Vercellati, C. Boschetti, W. Barcellini, P.G. Righetti, A. Zanella, Hum. Mutat. 30 (2009) 1292–1298.
[43] J. Denecke, T. Marquardt, Biochim. Biophys. Acta, PMID: 19150496 [Epub ahead of print].
[44] S. Hartwig, A. Czibere, J. Kotzka, W. Paßlack, R. Haas, J. Eckel, S. Lehr, Arch. Physiol. Biochem. 115 (2009) 155–160.