ELSEVIER

Pharmacoproteomics: a chess game on a protein field

Angelo D'Alessandro and Lello Zolla

Department of Environmental Sciences, Tuscia University, Largo dell'Università snc, 01100, Viterbo, Italy

The application of proteomics in the field of drug discovery development and the assessment of drug administration is known as pharmacoproteomics. As a branch of proteomics – perhaps the most promising and rapidly evolving field of the post-genomic era – pharmacoproteomics has inherited all the promises that pharmacogenomics has hitherto left unfulfilled. On the road to tailor-made drugs, whole protein profiles of healthy individuals have been progressively expanded, either qualitatively or quantitatively. In this review article, we provide general information about technical advancements in the field of proteomics (the pieces of this intriguing chess game) and show how this progress has furthered our understanding of biological systems. Pitfalls on the field of biomarker individuation and drug discovery and/or testing are also discussed.

Functional proteomics: the pharmacoproteomic promise

The application of proteomic technologies in the field of drug discovery development and assessment of drug administration is known as pharmacoproteomics [1]. Pharmacoproteomics directly stems from proteomics, the most rapidly evolving field since the end of the post-genomic era, and holds the potential to fulfill all the promises that pharmacogenomics has hitherto left unfulfilled. Although individualized medicine treatments still remain an ongoing objective, pharmacoproteomics seems to be more suited for providing valuable information in either the design or the toxicity assessment of new drugs because it sheds light on the effects of their interactions with the actual bioactive product, the protein. This is particularly evident when considering the field of biomarker discovery, in which mere genome-oriented or *in silico*-only approaches have so far failed to gather conclusive information, whereas the majority of disease diagnosis-related markers have been finally 'found in translation', to quote Lockhart and Walther [2]. Notwithstanding the big strides made by proteome-oriented approaches, only a handful of proteins are currently used in routine clinical diagnosis, and the rate of introduction of new protein tests approved by the United States Food and Drug Administration has paradoxically declined over the past two decades to fewer than one new protein diagnostic marker per year [3].

The performance of combinatorial chemistry in filling pharmaceutical pipelines has been lower than anticipated and the tide might be turning back to nature in the search for new drug candidates. Even though diversity-oriented synthesis is now producing molecules that are natural-product-like in terms of size and complexity, these molecules have not evolved to interact with biomolecules: in both cases, the mechanisms of action of these molecules are often unclear. Chemical and reverse chemical proteomics have recently contributed to speeding up the process of individuating new druggable targets and their small-molecule ligands, suggesting an alternative route to follow to overpass undesired drug market stagnancy [4].

New strategies have become necessary to tackle these difficulties. First, the whole protein profiles of healthy individuals have been progressively expanded, from a purely qualitative point of view. This is particularly true for plasma and blood components, the protein lists of which have been continuously improved during the past few years [5]. To this end, the introduction of new technologies enabling sample pre-fractionation, such as combinatorial ligand libraries [6], has allowed a whole hidden proteome to be unveiled. A paradigmatic example is the case of the red blood cell, in which the 98% of the whole cytosolic protein content is characterized by hemoglobin, and the remaining 2% has been

Corresponding author:. Zolla, L. (zolla@unitus.it)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.10.002

shown to harbor approximately 1578 unique proteins [7]. This amazing list of new proteins has been exploited to update the total map of erythrocyte proteins, from approximately 508 entries to 1989 distinct gene products [8]. In parallel, newly introduced technical caveats have endowed researchers with powerful tools – such as blue and clear native gel-based approaches – to explore the more hydrophobic protein fraction [9]. These approaches are relevant because they shed light on a considerable fraction of proteins that are difficult to separate and analyze with classical investigations, and they could also provide details about the formation of protein complexes, which is a recurring feature of enzymatic machineries in biological systems.

These results should be completed or accompanied by *in silico* modeling, prediction and evaluation of the likely effects of drug–protein interactions, or protein–protein interactions themselves, because the protein itself does not exist in the cell as an independent entity; rather, it takes part in intricate networks of interacting molecules [10]. It is worth underlining that, until recently, approaches to biomarker identification have sought to find single molecules indicative of normal or anomalous conditions (i.e. to distinguish healthy cells from diseased ones, such as progenitors or terminally committed mature cells). Indeed, the molecular complexity is an emerging property of living cells, and reductionist approaches have, so far, restrained an exhaustive understanding of the molecular processes. In this view, 'omics'-oriented approaches have been gaining momentum in the past two decades. The foremost example is perhaps proteomics, which studies the proteome, the cell-specific protein complement to the genome, and encompasses all the proteins that are expressed in a cell at the given time and under the given conditions (normal, stress, disease and tumor) in which the experiment is performed [1].

Second, new directions have been indicated and included in the ambitious agenda of the Human Proteome Organization (HUPO), in merit of the Plasma Proteome Project, which regroups 35 collaborating laboratories worldwide. In this view, quantitative data should be gathered, along with information about post-translational modifications, which proteins undergo through consolidated mechanisms of functioning regulation and/or control [11]. Therefore, if on the one hand the accumulation of qualitative information about which proteins are expressed in healthy tissues is fundamental, on the other hand it is also pivotal to determine the fluctuations of quantities and the variations in patterns of post-translational modifications of these proteins in specific tissues under healthy and pathological conditions, before or upon drug assumption.

Being this the chessboard, in this Review article we are briefly providing general information about technical advancements in the field of proteomics (the pieces of this intriguing chess game; Fig. 1) and showing how these advancements have furthered our understanding of biological systems. Pitfalls in the field of biomarker individuation and drug discovery and testing are also discussed.

The pawn: qualitative profiles – who, where and when

The first step of proteomics analyses is the determination of the specific pattern of protein expression of a given tissue. Tissue proteomic profiling becomes fundamental when it comes to comparing healthy and diseased patients, to determine markers that could be adopted for early diagnoses. In parallel, protein fingerprints of healthy tissues could vary upon drug assumption, thus representing a valuable tool in the assessment of new drug efficiency and effectiveness or to evaluate likely drug targets [12].

The very first phase of the proteomic era has been begun by the need to build up a knowledge body of qualitative databases.

Blood plasma: technical caveats

In this respect, relevant results have been obtained in the field of transfusion medicine, as it has been reviewed recently by several authors [5,13]. Indeed, blood components particularly lend themselves to protein-oriented approaches because the enucleated nature of red blood cells and platelets nullifies transcript-oriented analyses. Nonetheless, the analysis of the protein fractions of blood and blood components still represents a challenging task. The high dynamic range spread of blood and blood component proteins hampers a unitary and comprehensive view of both high- and low-abundant species, including proteins involved in the most different activities (coagulation, transport, immune system and cell signaling) and protein byproduct from cellular damage of other tissues. Blood samples can be taken at a particular point in time with little burden on patients, and the constituents of the blood samples could reflect a developing or existing illness because tissue-specific proteins might be released into the blood stream from the damaged or dead cells [14]. The latter consideration suggests new scenarios because the qualitative study of minor protein populations in blood components not only is tied to transfusion medicine but could also represent a virgin forest in the field of new biomarker discovery. In this view, blood could constitute a reservoir of proteins suggestive of specific pathological conditions.

Nonetheless, blood and blood components are characterized by an extreme dynamic range of proteins, the concentrations of which span more than ten orders of magnitude, from picogram to milligram quantities per milliliter in the case of plasma. Several high-abundance proteins, such as albumin, typically constitute greater than 90% of total protein mass. Therefore, the detection of lower abundance proteins, which presumably represents the biologically interesting population, is interfered with by the dominant proteins. Sample preparation and pre-fractionation become pivotal when tackling the complexity of these samples, through the removal of high-abundant species through different approaches.

The bishop: sample fractionation

Because of the high complexity of biological samples (especially of blood and blood components), several strategies have been proposed for cutting through the dynamic range of protein concentrations and variability spread. Basically, these methods rely on biochemical and biophysical properties of protein/peptide molecules, *viz.*, molecular weight, density, hydrophobicity, surface change and isoelectric point, and inter-molecular affinity (e.g. protein–protein complex formation). Although samples could be split through preliminary electrophoresis pre-fractionation, most of the methods are based on the removal of high-abundant species. Immunoaffinity depletion [14], centrifugal ultrafiltration [15] and combinatorial ligand libraries [6] are three of the most widely diffused strategies to tackle dynamic concentration range-related analytical issues.

FIGURE 1

Pharmacoproteomics as a chess game. Proteomic strategies and chess game pieces are briefly enlisted and described.

Resin-based (e.g. Cibacron blue, blue dye) and antibody-based (e.g. IgY directed towards multiple high-abundant protein species) depletion are probably the most widely diffused pre-fractionation methods in biological analyses. Commercial kits enable the removal of the 6 (albumin, transferrin, IgG, IgA, haptoglobin and antitrypsin) [16], 20 [17] and 89 [18] most abundant protein species, thus cutting the analytical noise by more than 97%. Alternatively, avian immunoglobulin yolk (IgY) kits are available, displaying higher avidity and less cross-reactivity with heterologous human proteins.

High-abundant protein removal has its drawback as well, owing to the so-called 'albumin sponge effect', which implies that up to 210 different minor protein species (potential candidate biomarkers) are found to be associated with the six most abundant plasma proteins (in particular with albumin) [19]. For example, the removal of albumin using a resin causes a considerable loss of several cytokines. The sponge effect could be tackled through salt-out preparation and molecular sieve filtration [20]. It might also be possible to design a non-protein binding resin or membrane containing small peptides to selectively remove albumin, but not albumin-bound proteins, from plasma [21]. Nonetheless, immunodepletion-based technology still has some considerable technical issues, which include short life span (<200 runs), progressive reduction of antibody binding capability towards specific proteins or altered specificity through high-molecular-weight compound-induced absorption through hydrophobic interactions [14], altered proportion of quantitative outputs and small sample loading capacity (25–100 ml), not to mention the elevated costs for columns and buffers.

Combinatorial hexapeptide ligand libraries have emerged recently as a valuable alternative to immunodepletion [6]. The combinatorial peptide ligand library is a mixture of porous beads on which hexapeptides are chemically attached. Each bead carries 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. By 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 reducing the relative concentration of abundant species [22]. Indeed, the combinatorial hexapeptide ligand library approach has contributed to dramatically delving into the complexity of a wide series of biological samples. 2-DE gels upon treatment with combinatorial peptide ligand libraries revealed a far greater number of protein species, as in the case of urine (from 184 to 385 [23]), plasma (from 115 to 790 [24]) and platelets (from 197 to 435 [25]), as well as in other biological fluids, such as milk fractions or in egg yolk and white [26]. A reduced percentage of the proteins that are visible with common methods gets lost; however, 7% in the case of urine, 5% for red blood cells and 13% for platelets [24]. The continuous expansion of the proteomes of biological samples holds that a wider basis is available for statistical and *in silico* analyses to determine molecular species of interest for human health (novel candidate biomarkers, drug-like molecules, anti-microbial peptides, etc.).

Organellar proteomics and subcellular fractionation: A further level of complexity is introduced by the precisely controlled spatial organization of these proteins, namely where they are actually expressed, not only in terms of tissues or cells but also in terms of subcellular compartment [27]. A simple hierarchy of cellular organization would include cellular compartments, such as the cytosol and nucleus, membrane-enclosed organelles, and large or small multiprotein complexes. Large-scale approaches to organelle proteomes have been increasingly attracting a great deal of interest as Taylor et al. have recently reviewed in the literature [28]. After an organelle has been biochemically purified, it can be analyzed through proteomic approaches: in this respect, 2-DE has been almost completely substituted by approaches involving preliminary 1D-SDS-PAGE, enzymatic digestion of band of interest, chromatographic separation of the resulting peptides and subsequent identification via mass spectrometry [29]. Recently, shotgun proteomics was used to demonstrate the implication of protease inhibition and complement activation in the anti-inflammatory properties of high-density lipoprotein [30]. This workflow enables optimal protein identification, but a validation step is still required to assess whether the identified protein is a genuine component of a specific organelle. Moreover, it should be considered that the protein constituents of most organelles are in constant exchange with the rest of the cell. Several organelles – particularly nuclear subdomains – contain both resident and transient proteins. Perturbations and temporal changes in organelle proteomes could be investigated through stable isotope of amino acid in culture (SILAC) upon blockade of transcription to stabilize the cellular subproteomes at different times (up to nine).

The knight: sample preparation

Sample preparation and fractionation technologies are two of the most crucial processes to delve into the complexity of solubilized samples. There are, however, considerable limitations in currently available proteomic technologies: none of them allow for the analysis of the entire proteome in a simple step. This is due to the large number of peptides and to the wide concentration dynamic range of the proteomes in clinical blood samples. As a result, it is often observed that one sample preparation approach biases the proteomic outcome towards a slightly different output; therefore, an ideal proteomic analysis would tend to reduce sample preparation steps [31].

Plasma proteins: sample preparation – challenge ups its ante. An adequate analysis generally includes the following steps: (i) sampling, in which the sample is a good statistical representation of the investigated population; (ii) specimen preservation, during which the sample is expected to be kept stable until the analysis is completed; (iii) appropriate sample handling (fractionation); and (iv) statistical analysis and bioinformatics data treatment [31]. Ideally, sample preparation should be as simple as possible to reduce time and avoid the introduction of steps that could lead to sample loss. Studies from literature seem to attribute the most relevant role in proteomic standardization to sampling strategies, whereas handling and storage conditions apparently provoke relatively minor effects [32]. Indeed, minimal changes were observed in the samples stored at room temperature within the first four hours or six hours, whereas noticeable changes were found after eight hours, especially for peaks in the m/z range at 3000 [33], and were more pronounced after 24 h [34].

Nonetheless, these parameters need adequate standardization as well, in particular for plasma- and blood component-derived proteins, to have comparable results between different laboratories [33]. Standardization has also become an urgent issue for the HUPO, which recently started an international standardization program in proteomics (the Proteomics Standardization Initiative) [11], with the declared intent of defining community standards for data representation in proteomics, which should ease data comparison, exchange and verification. In parallel, the minimum information about a proteomics experiment (MIAPE) initiative has been started to define community standards for data production and representation in proteomics, which should reduce technical variability and ease data comparison, exchange and verification [35].

Other than storage or handling-related changes, proteomic analyses are mainly affected by sample preparation procedures, which include preliminary protein solubilization steps [31].

The rooks and the queen: membrane proteins, protein complexes and protein interactions

First rook: membrane proteins. Proteins in biological samples are often insoluble because in their native state they could be found in the form of molecular complexes of associated proteins or in membranes. Solubilization steps vary depending on which kind of protein population is addressed, whether protein complexes or membrane-associated proteins.

When promoting solubilization of protein complexes, it is necessary to break interactions involved in protein aggregation (e.g. disulfide and hydrogen bonds, van der Waals forces, ionic and hydrophobic interactions), which enables disruption of proteins

into a solution of individual polypeptides, thereby promoting their solubilization. Sample solubilization can be improved by agitation or ultrasonification, but an increase in temperature must be avoided. The selection of the appropriate solubilization protocol and buffers has especially been facilitated by the availability of commercial kits [36], although it is somewhat more expensive than routine reagent methods. Ionic detergents are strong solubilizing agents that produce protein denaturation. High concentrations of chaotropes such as urea (5–9 M) and thiourea (2 M) help to increase the overall number of solubilized proteins by disrupting hydrogen bonds, whereas charged chaotropic agents such as guanidine hydrochloride should be avoided because they are not compatible with isoelectrofocusing (IEF).

Detergents are used to solubilize proteins and consist of a polar head group and a hydrophobic tail, enabling them to solubilize membrane proteins mimicking the lipid environment. The gentleness of detergent is mainly determined by the size of its polar head group and the length of its hydrophobic acyl tail [37]. Detergents could be used at high, medium and low concentrations. When targeting separated proteins, detergents are used within a 1–4% concentration range, to prevent hydrophobic interactions. It is important to know whether the detergent should be removed and, if so, how easy it is to remove it – for example, by dialysis. In sample preparation for 2-DE, only neutral (octylglucoside, dodecyl maltoside and Triton X-100) or zwitterionic (3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propane sulfonate, CHAPSO, SB 3-10, SB 3-12 and ABS-14) detergents are used, owing to their compatibility with the separation mechanisms. The anionic detergent sodium dodecyl sulfate (SDS) improves solubilization but interferes with the first electrophoretic dimension separation and must be removed if present in the preparation (e.g. ETTanTM 2D Clean Up kit, GE Health Care; ProteoSpin^{TM} Detergent Clean UP micro and maxi kits, Norgen Biotek Corporation). To protect proteins from protease digestion, inhibitors should also be included during sample preparation phases [38].

Solubilization protocols should be adapted when handling hydrophobic proteins, such as membrane proteins. Membrane proteins represent a large population of the proteome in the form of receptors, transporters, channels and a variety of cellular mechanisms. The importance of membrane proteins is highlighted by the fact that approximately one-third of all genes in various organisms code for this class of proteins [39]. More than two-thirds of all medications exert their effects through membrane proteins [40], which makes them a major target of pharmacological interest. However, membrane proteins are still under-identified and underrepresented during whole-cell proteome analysis. Although membrane proteins with up to 12 transmembrane α-helices have been resolved and identified by 2-DE-MS [41], most membrane proteins have been resistant to this approach.

Membrane proteins are often enriched by ultracentrifugation in sucrose gradient, lectin affinity chromatography in combination with centrifugation, silica beads or biotinylation and interaction with immobilized streptavidin [42]. Solubilization of membrane proteins is achieved with different detergents, such as combinations of chloroform and methanol, which were used to extract hydrophobic chloroplast membrane proteins. Alternatively, aqueous two-phase systems (which employ detergent DDM, Triton X-114 or PEG for the selective binding of one or more proteins of interest to one of the incompatible aqueous phases) could be used.

Second rook: protein complexes. Proteins rarely function in isolation. Rather, they are organized in functional units that are different in size, number of interacting partners and stability, ranging from huge stable ribosomes or nuclear pore domains to small and transient signal transduction complexes.

Studying these multiprotein complexes and microdomains provides information about the spatio-temporal organization of signal transduction or metabolic processes within a cell. Moreover, most of these protein complexes involve membrane proteins. However, a major part of this information is lost when cells are lysed and proteins digested before analysis. Isolating protein from complexes enables reduced complexity and eases the identification of low copy number proteins and their specific particular functions [43], although most of the sample pre-fractionation strategies (see the section 'The bishop: sample fractionation') tend to disrupt protein complexes by negatively influencing native protein–protein interactions. Protein complexes and hydrophobic protein species can be investigated through different approaches, including affinity-based methods, recombinant pull-downs, liquid chromatography, blue native gel electrophoresis (BN-PAGE), 2-DE/LC/CE and FFE methods, followed by MS analysis [41]. Protein denaturing approaches (IEF, SDS-PAGE, 16-BAC-PAGE and CTAB-PAGE) are not suited to tackle protein complexes [41]. Nevertheless, BN-PAGE is perhaps the gel-based criterion of choice when addressing protein complexes and membrane proteins. BN-PAGE was introduced by Schagger and Von Jagow almost 20 years ago [44] and exploits Comassie G-250 to partially charge proteins to increase their electrophoretic mobility. Native gel-based approaches target protein complexes in their native form, as an alternative to immune co-precipitation [45], against which they enable separation without antibodies at the expense of detergent-labile interacting protein loss.

Gel-free platforms particularly lend themselves to shotgun proteomic approaches, such as MudPIT, and minimize the problem of insolubility encountered in membrane protein studies [46].

The queen: protein–protein interactions

Protein array information can be exploited to obtain complex maps of proteins involved in disease-specific molecular signatures and could, therefore, be translated to *in silico* data handling through elaborations of protein–protein interaction maps [47]. Biomolecule interactions are also relevant in that they might elucidate the efficiency and effectiveness of a targeted drug treatment. For example, the extent of drug binding to plasma proteins, determined by measuring the free active fraction, has a notable effect on the pharmacokinetics and pharmacodynamics of a drug. To this end, a vast array of different methods have been developed, *viz.*, equilibrium dialysis, ultrafiltration, parallel artificial membrane permeability assay, high-performance affinity chromatography/zonal elution approach, high-performance affinity chromatography/frontal analysis approach, affinity capillary electrophoresis, capillary electrophoresis/frontal analysis, spectroscopic assays, isothermal titration calorimetry, competition studies, titration studies, differential scanning calorimetry, surface plasmon resonance-based assays and SILAC [47]. Vice versa, it has been hypothesized that targeted drug treatments should address target protein biomarkers,

which are characterized by specific properties making them eligible interactors of the designed drug ('drug target-likeness') [48]. In this respect, *in silico* modeling becomes pivotal for optimal design of drugs against, for example, specific subclasses of receptors [49].

In parallel, progresses have been made recently in the field of protein–protein interaction mapping and elaboration of networks, pathways and enrichment of gene ontologies for biological and molecular functions, especially upon the introduction of user-friendly software and updated databases [50] (Fig. 2). The ultimate goal of this kind of investigation is to use the software output to focus on specific targets as early biomarkers of pathological conditions or druggable-like targets for medical treatment.

In parallel, the *in silico* approach could be useful in creating models or elaborating data from experimental observations about changes in pathways and networks upon administration of specific drugs. These methods might reveal drug effects on pathways, which represents the cornerstone of identifying mechanisms of drugs' efficacy.

The king: quantitative profiles

Pre-analytical control

Although a wealth of studies are regularly being published about successful proteomic approaches to biological marker (biomarker) discovery, the discovery of new diagnostic biomarkers lags behind because of variability at every step in proteomics studies (e.g. assembly of a cohort of patients, sample preparation and the nature of body fluids, and selection of a profiling method and uniform protocols for data analysis) [51]. Most of this gap is due to

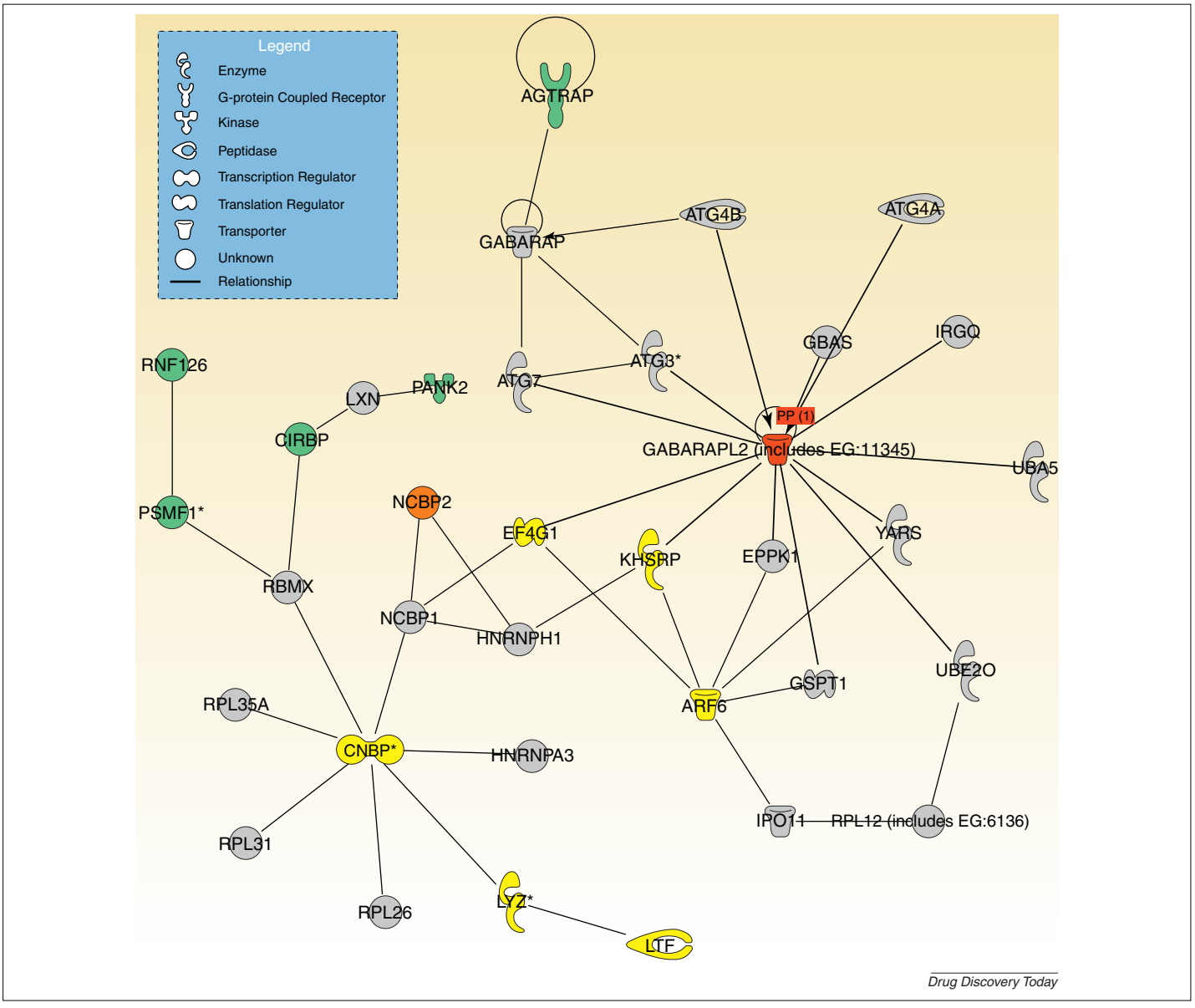

FIGURE 2

Network analysis through Ingenuity Pathway Analysis of an experimental dataset. The software graphs proteins as nodes and interactions as connecting hubs (edges). Nodes could vary in color and shape to reflect experimental quantitative fluctuations or indicate different biological functions. Red nodes are overexpressed among two groups under comparison. Highly connected nodes (high-degree) represent the most biologically relevant proteins because they function as regulatory hubs for multiple pathways. Edges could be represented by continuous or interrupted lines, to indicate direct or indirect interactions between two proteins.

the lack of proper quantitative results on large cohorts of patients. As a result, most of the currently available biomarker discovery studies lack the statistical strength for the identification of high-confidence biomarkers and are thus more suited to tailor-made studies, although they lack the statistical strength of large-scale studies [52]. Often, the validation step that follows the discovery phase does not reach desired levels of sensitivity and specificity or reproducibility between laboratories [51]. Identification of the importance of preanalytical factors has implications for the use of large sample banks [53]. Such banks enable many prospective studies, which otherwise would require years for sample acquisition, to be carried out in a timely fashion, but sample banks vary in their adherence to consistent sample processing protocols over time. Many proteins, however, can be unaffected by many of the preanalytical variables. Therefore, any study using or proposing to use such banks should not be instantly condemned, but the way in which the banked samples are used should be examined.

A possible, although imperfect, solution to this issue is to restrict the initial phase of biomarker discovery to specimens collected by rigorous adherence to banking protocols.

To overcome this hurdle, multi-group efforts are necessary to facilitate the generation of sufficient sample sizes. This is contingent on the ability to collate and cross-compare data from different studies, which will require the use of a common metric or standard.

Checkmate: biomarker discovery and quantitation

The biomarker field has generated various examples of individual biomarkers that can be used as indicators of disease pathophysiology, such as blood pressure, cholesterol levels and viral load in HIV [54]. The complexity of the underlying pathophysiological pathways and interactions makes it desirable to identify shortcuts through the proteome to associate meaningful biomarker-like signatures to specific protein profiles, rather than relying on independent biomolecules alone. The identification of a pattern or profile of several biomarkers (representing, e.g. a combination of genes, proteins, organic molecules, metabolites and/or a physiological response) representative of a given condition might bring a new dimension to disease diagnosis, classification and intervention and the assessment of therapeutic responses.

Nonetheless, to be eligible as a valuable biomarker candidate, a protein should be detectable early and quantifiable over disease progression, and its co-occurrence with the disease should be statistically significant in a robust cohort of patients: although most scientific efforts have been put forward to address the first and last concerns, quantification has become a pivotal node in biomarker discovery and validation in recent years, especially in the field of human cancer research. To compare the abundance of proteins in different samples, several quantitative approaches have been developed (such as SILAC, isotope-coded affinity tags, exponentially modified protein abundance index and subtractive proteomic strategies), as Veenstra brilliantly reviewed [55]. Soon enough, quantitation strategies will enable biomarker discovery to meet quality criteria of clinical laboratory medicine.

In clinical laboratory medicine

Clinical proteomics consists of discovery proteomics and measurement (quantitative) proteomics. Our current main focus is still discovery proteomics, clinical proteomics will undoubtedly be a major contributor to laboratory medicine in the near future.

The potential clinical benefits for disease-specific biomarkers (to be tested for in clinical routine practice) include a more rapid and accurate disease diagnosis and a potential reduction in size and duration of clinical drug trials, which would speed up drug development. For example, proteomics in urine analysis could help identify and quantify proteins excreted in urine, which are not only key indicators of diseases associated with renal function but also indicators of the overall health of individuals.

The application of biomarkers in the clinical arena of motor neuron disease should determine both whether a drug hits its proposed target and whether the drug alters the course of disease. It is interesting to note that, despite blood lacking direct contact with brain, increasing evidence suggests that there is a blood protein signature, and possibly a transcript signature, that might act to increase confidence in diagnosis and be used to predict progression in either disease or prodromal states. It might also be used to monitor disease progression [56]. These optimistic results come partly from candidate protein studies, from which it emerged that amyloid-beta measures might have value in prediction, and those studies of inflammatory markers that consistently show change in Alzheimer's disease [56].

In theranostics

Non-invasive biomarker strategies based on imaging technologies – including magnetic resonance imaging, single-photon emission computed tomography, positron emission tomography and other techniques – are also progressing quickly in various therapeutic areas [57].

State-of-the-art applications of proteomics in the field of biomarker discovery involve the elaboration of specific signatures in 2DE maps, quantitative approaches [58], the individuation of marker-like post-translational modifications [59] and, recently, the measurement of varying metabolic profiles [60].

Complementary strategies are now adopted to focus on disease-relevant converging pathways as potential therapeutic intervention points and are accompanied by the individuation of downstream biomarkers, which enable the tracking of drug targeting and seem to correlate with disease mitigation. When putting the pieces together, one is able to envision that a companion diagnostic will be codeveloped along the therapeutic compound. This 'theranostic' approach is perfectly positioned to align with the emerging trend toward personalized medicine [61]. In the quest for tailor-made drugs, proteomics has been contributing with previously unexpected results, involving, for example, gender-specific differences in human serum, which could be fundamental to building gender-specific protein biomarker databases to be exploited for targeted treatments.

Protein chips

Once individuated, however, it is very unlikely that biomarkers will be monitored with 2-DE or mass spectrometry, which are routine techniques in the academic setting but do not lend themselves to clinical routine practice on large cohorts of samples. Further development of protein arrays might address this issue [51].

A faster approach to biomarker monitoring, using the integration of genomic and proteomic techniques, is based on the pro-

duction of protein chips. In this approach, cDNAs encoding tagged proteins are expressed and proteins are isolated and printed on a slide. These slides are then used to examine protein profiling, protein–protein interactions and antibody profiling [62]. These protein array techniques [63] can help not only to detect potential novel biomarkers but also to generate a greater understanding of the signaling pathways associated with the printed proteins.

Protein-detecting arrays use a wide variety of capture agents (antibodies, fusion proteins, DNA/RNA aptamers, synthetic peptides, carbohydrates and small molecules) immobilized at high spatial density on a solid surface. Each capture agent binds selectively to its target protein in a complex mixture, such as serum or cell lysate samples. Captured proteins are subsequently detected and quantified in a high-throughput manner, with minimal sample consumption. Furthermore, the first multiplex immunoassay systems have been cleared by the US Food and Drug Administration, signaling recognition of the usefulness of miniaturized and parallelized array technology for protein detection in predictive and/or early diagnosis. Although genetic tests still predominate, with further development protein-based diagnosis will become common in clinical use within a few years [64].

Concluding remarks

One square at a time, qualitative protein profiles of healthy and diseased tissues have been thoroughly investigated. Sample preparation and fractionation techniques have been gradually optimized to cut through the biological complexity of blood components, making it possible to identify protein signatures of diagnostic value and clinical importance. Alternative strategies have been created to target either protein biomarkers (often very low abundance protein) or drug-target-like molecules (often hydrophobic proteins at the membrane level) [65], both needing thorough clinical validation on large cohorts of patients before their widespread diffusion. The real deal is the limited biological variability tested in discovery studies, the current lack of quantitative information and, thus, the need for a proper validation and the absence of quality criteria in alignment to standard clinical practice. These objectives represent the very heart of the ambitious agenda and need to be accomplished in the next few years.

The growing need for quantitative information is linked to the opportunity to better understand biological systems as a whole, through the integration of quantitative proteomics, metabolomics and transcriptomics auspicated by systems biology. Whether tailor-made treatments are still far from being a reality, pharmacoproteomics and *in silico* elaborations have contributed to make this chess game more challenging and intriguing than ever before. As huge amounts of data are being accumulated, the fulfillment of the pharmacoproteomic promises is rapidly approaching; however, the run towards the checkmate is not yet over.

Acknowledgements

ADA and LZ were supported by founds from the Italian National Blood Centre (Centro Nazionale Sangue – Istituto Superiore di Sanità – Rome, Italy).

References

1 Zolla, L. (2008) Proteomics studies reveal important information on small molecule therapeutics: a case study on plasma proteins. *Drug Discov. Today*. 13 (Suppl. 23–24), 1042–1051

2 Lockhart, B.P. and Walther, B. (2009) Biomarkers: 'found in translation'. *Med. Sci. (Paris)* 25, 423–430

3 Anderson, N.L. and Anderson, N.G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867

4 Piggott, A.M. and Karuso, P. (2004) Quality, not quantity: the role of natural products and chemical proteomics in modern drug discovery. *Comb. Chem. High Throughput Screen.* 7, 607–630

5 Liumbruno, G. *et al.* (2009) Blood-related proteomics. *J. Proteomics* 73, 483–507

6 Righetti, P.G. and Boschetti, E. (2009) Blood proteomics and the dynamic range: some light at the end of the tunnel? *J. Proteomics* 73, 627–628

7 Roux-Dalvai, F. *et al.* (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

8 D'Alessandro, A. *et al.* (2010) The red blood cell proteome and interactome: an update. *J. Proteome Res.* 9, 144–163

9 Braun, R.J. *et al.* (2007) Two-dimensional electrophoresis of membrane proteins. *Anal. Bioanal. Chem.* 389, 1033–1045

10 Lee, S. *et al.* (2009) Structural interactomics: informatics approaches to aid the interpretation of genetic variation and the development of novel therapeutics. *Mol. Biosyst.* 5, 1456–1472

11 Omenn, G.S. *et al.* (2009) 7(th) HUPO World Congress of Proteomics: launching the second phase of the HUPOPlasma Proteome Project (PPP-2) 16–20 August 2008, Amsterdam, The Netherlands. *Proteomics* 9, 4–6

12 Chanda, S.K. and Caldwell, J.S. (2003) Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov. Today* 8, 168–174

13 Thiele, T. *et al.* (2007) Proteomics of blood-based therapeutics. *BioDrugs* 21, 179–193

14 Ichibangase, T. *et al.* (2009) Limitation of immunoaffinity column for the removal of abundant proteins from plasma in quantitative plasma proteomics. *Biomed. Chromatogr.* 23, 480–487

15 Greening, D.W. and Simpson, R.J. (2010) A centrifugal ultrafiltration strategy for isolating the low-molecular weight (≤25 K) component of human plasma proteome. *J. Proteomics* 73, 637–648

16 Huang, L. *et al.* (2005) Immunoaffinity separation of plasma proteins by IgY microbeads: meeting the needs of proteomic sample preparation and analysis. *Proteomics* 5, 3314–3328

17 Lin, B. *et al.* (2009) Deep depletion of abundant serum proteins reveals low-abundant proteins as potential biomarkers for human ovarian cancer. *Proteomics Clin. Appl.* 3, 853–861

18 Qian, W.J. *et al.* (2006) Advances and challenges in liquid chromatography–mass spectrometry-based proteomics profiling for clinical applications. *Mol. Cell. Proteomics* 5, 1727–1744

19 Lee, H.J. *et al.* (2006) Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics. *Curr. Opin. Chem. Biol.* 10, 42–49

20 Hey, J. *et al.* (2008) Fractionation of complex protein mixtures by liquid-phase isoelectric focusing. *Methods Mol. Biol.* 424, 225–239

21 Jurgens, M. and Schrader, M. (2002) Peptidomic approaches in proteomic research. *Curr. Opin. Mol. Ther.* 4, 236–241

22 Righetti, P.G. and Boschetti, E. (2008) The ProteoMiner and the FortyNiners: searching for gold nuggets in the proteomic arena. *Mass Spectrom. Rev.* 27, 596–608

23 Castagna, A. *et al.* (2005) Exploring the hidden human urinary proteome via ligand library beads. *J. Proteome Res.* 4, 1917–1930

24 Sennels, L. *et al.* (2007) Proteomic analysis of human blood serum using peptide library beads. *J. Proteome Res.* 6, 4055–4062

25 D'Amato, A. *et al.* (2009) In-depth exploration of cow's whey proteome via combinatorial peptide ligand libraries. *J. Proteome Res.* 8, 3925–3936

26 D'Alessandro, A. *et al.* (2010) The egg white and yolk interactomes as gleaned from extensive proteomic data. *J. Proteomics* 73, 1028–1042

27 Andersen, J.S. and Mann, M. (2006) Organellar proteomics: turning inventories into insights. *EMBO Rep.* 7, 874–879

28 Taylor, S.W. *et al.* (2003) Global organellar proteomics. *Trends Biotechnol.* 21, 82–88

29 Yates, J.R. *et al.* (2005) Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell Biol.* 6, 702–714

30 Vaisar, T. *et al.* (2007) Shotgun proteomics implicates protease inhibition and complement activation in the antiinflammatory properties of HDL. *J. Clin. Invest.* 117, 746–756
31 Ahmed, F.E. (2009) Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. *J. Sep. Sci.* 32, 771–798
32 Hsieh, S.Y. *et al.* (2006) Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. *Proteomics* 6, 3189–3198
33 Rai, A.J. *et al.* (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 5, 3262–3277
34 Marshall, J. *et al.* (2003) Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J. Proteome Res.* 2, 361–372
35 Taylor, C.F. *et al.* (2007) The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 25, 887–893
36 Cañas, B. *et al.* (2007) Trends in sample preparation for classical and second generation proteomics. *J. Chromatogr. A* 1153, 235–258
37 Reisinger, V. and Eichacker, L.A. (2008) Solubilization of membrane protein complexes for blue native PAGE. *J. Proteomics* 71, 277–283
38 Tan, S. *et al.* (2008) Membrane proteins and membrane proteomics. *Proteomics* 8, 3924–3932
39 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
40 Plewczynski, D. and Rychlewski, L. (2009) Meta-basic estimates the size of druggable human genome. *J. Mol. Model.* 15, 695–699
41 Helbig, A.O. *et al.* (2010) Exploring the membrane proteome – challenges and analytical strategies. *J. Proteomics* 73, 868–878
42 Komatsu, S. *et al.* (2007) The proteomics of plant cell membranes. *J. Exp. Bot.* 58, 103–112
43 Pasquali, C. *et al.* (1999) Subcellular fractionation, electromigrationanalysis and mapping of organelles. *J. Chromatogr. B: Biomed. Sci. Appl.* 722, 89–102
44 Schagger, H. and von Jagow, G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* 199, 223–231
45 Blonder, J. *et al.* (2002) Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography–tandem mass spectrometry. *J. Proteome Res.* 1, 351–360
46 Han, D.K. *et al.* (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 19, 946–951
47 Ong, S.E. *et al.* (2009) Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4617–4622
48 Xu, H. *et al.* (2007) Learning the drug target-likeness of a protein. *Proteomics* 7, 4255–4263
49 Sela, I. *et al.* (2010) G protein coupled receptors – *in silico* drug discovery and design. *Curr. Top. Med. Chem.* 10, 638–656
50 Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504
51 Silberring, J. and Ciborowski, P. (2010) Biomarker discovery and clinical proteomics. *Trends Analyt. Chem.* 29, 128
52 Chao, T.C. *et al.* (2010) Towards proteome standards: the use of absolute quantitation in high-throughput biomarker discovery. *J. Proteomics* 73, 1641–1646
53 Banks, R.E. (2008) Preanalytical influences in clinical proteomic studies: raising awareness of fundamental issues in sample banking. *Clin. Chem.* 54, 6–7
54 Ilyin, S.E. *et al.* (2004) Biomarker discovery and validation: technologies and integrative approaches. *Trends Biotechnol.* 22, 411–416
55 Veenstra, T.D. (2007) Global and targeted quantitative proteomics for biomarker discovery. *J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci.* 847, 3–11
56 Thambisetty, M. and Lovestone, S. (2010) Blood-based biomarkers of Alzheimer's disease: challenging but feasible. *Biomark. Med.* 4, 65–79
57 Rudin, M. and Weissleder, R. (2003) Molecular imaging in drug discovery and development. *Nat. Rev. Drug Discov.* 2, 123–131
58 Liu, Y.F. *et al.* (2009) Quantitative proteome analysis reveals annexin A3 as a novel biomarker in lung adenocarcinoma. *J. Pathol.* 217, 54–64
59 Papini, A.M. (2009) The use of post-translationally modified peptides for detection of biomarkers of immune-mediated diseases. *J. Pept. Sci.* 15, 621–628
60 Wibom, C. *et al.* (2010) Metabolomic patterns in glioblastoma and changes during radiotherapy: a clinical microdialysis study. *J. Proteome Res.* 9, 2909–2919
61 Zhang, Z. *et al.* (2010) Systems biology and theranostic approach to drug discovery and development to treat traumatic brain injury. *Methods Mol. Biol.* 662, 317–329
62 Tomizaki, K.Y. *et al.* (2010) Protein–protein interactions and selection: array-based techniques for screening disease-associated biomarkers in predictive/early diagnosis. *FEBS J.* 277, 1996–2005
63 Miike, K. *et al.* (2010) Proteome profiling reveals gender differences in the composition of human serum. *Proteomics* 10, 2678–2691
64 Tomizaki, K.Y. *et al.* (2010) Protein–protein interactions and selection:array-based techniques for screening disease-associated biomarkers in predictive/early diagnosis. *FEBS J.* 277, 1996–2005
65 Lai, Z.W. *et al.* (2009) Membrane proteomics: the development of diagnostics based on protein shedding. *Curr. Opin. Mol. Ther.* 11, 623–631