Asymmetric Proteome Equalization of the Skeletal Muscle Proteome Using a Combinatorial Hexapeptide Library

Jenny Rivers¹, Chris Hughes², Thérèse McKenna², Yvonne Woolerton¹, Johannes P. C. Vissers², James I. Langridge², Robert J. Beynon¹*

¹Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom, ²Waters Corporation MS Technologies Centre, Atlas Park, Wythenshawe, Manchester, United Kingdom

Abstract

Immovilized combinatorial peptide libraries have been advocated as a strategy for equalization of the dynamic range of a typical proteome. The technology has been applied predominantly to blood plasma and other biological fluids such as urine, but has not been used extensively to address the issue of dynamic range in tissue samples. Here, we have applied the combinatorial library approach to the equalization of a tissue where there is also a dramatic asymmetry in the range of abundances of proteins; namely, the soluble fraction of skeletal muscle. We have applied QconCAT and label-free methodologies to the quantification of the proteins that bind to the beads as the loading is progressively increased. Although some equalization is achieved, and the most abundant proteins no longer dominate the proteome analysis, at high protein loadings a new asymmetry of protein expression is reached, consistent with the formation of complex assemblies of heat shock proteins, cytoskeletal elements and other proteins on the beads. Loading at different ionic strength values leads to capture of different subpopulations of proteins, but does not completely eliminate the bias in protein accumulation. These assemblies may impair the broader utility of combinatorial library approaches to the equalization of tissue proteomes. However, the asymmetry in equalization is manifest at either low and high ionic strength values but manipulation of the solvent conditions may extend the capacity of the method.

Introduction

Although the capability of mass spectrometry based proteomics has been greatly enhanced, the challenge of dynamic range is not fully solved. Individual proteins in a proteome are expressed at dramatically different levels, and the low abundance, or ‘deep’ proteome remains elusive. In blood plasma, the dynamic range of protein abundance may span as much as 12 orders of magnitude [1,2] and even in simple cellular systems such as yeast cytosol, expression may range over four orders of magnitude [3,4]. The most common approach to global proteome analysis is based on LC-MS/MS of tryptic peptides. The dynamic range of a typical LC-MS/MS analysis is such that it is not feasible to accommodate the range of protein expression levels in a typical tissue preparation. In particular, the low abundance proteins are difficult to analyze because of limited instrument sensitivity and ‘crowding’ of the analyte stream by high abundance peptides that trigger data dependent acquisition and which may cause ion suppression. Data independent data acquisition can ease this problem, but not completely.

Prefractionation and enrichment of subclasses of proteins such as those decorated with specific post-translational modifications can give selective enrichment [5], but can introduce the possibility of higher variance due to sample preparation. More commonly and generically, two strategies of selective proteome manipulation have been employed. First, depletion strategies are subtractive methods that use affinity methods, usually based on antigen:antibody interaction for selective removal of abundant proteins from a sample [6,7]. Almost without exception, this approach has been applied to human blood plasma [8–10] and has the goal of removal of major protein species that contribute the most intense ions in an LC-MS/MS analysis. Whilst such approaches can deplete abundant proteins (provided appropriate high specificity antibodies are available) they are not capable of enriching trace
proteins, and under certain conditions can be non-specific in removing selected high abundance proteins [11].

The second common approach to reduction of the dynamic range of a complex mixture of proteins is that of sample equalization [12–20]. This method uses combinatorial peptide libraries to generate a complex set of bead-immobilized ligands, each of which is able to bind, with variable affinity, a small subset of proteins in the analyte. A library of linear hexapeptides based on 20 naturally occurring amino acids has the potential to create a library of 64 million different ligands. Because each ligand is present at low abundance, the outcome is that high abundance proteins are bound to the beads at a finite capacity, and provided that adequate starting material is applied to the beads, low abundance proteins will eventually bind a similar number of beads, the outcome of which is to equalize the protein mixture attached to the beads.

Proteome equalization has predominantly been applied to analysis of human blood plasma, but lesserly to urine [12,13] and serum [15,21] as well as non-blood samples including cell and tissue lysates [22–27]. However, the emphasis on secreted biofluids, and in particular blood plasma has tended to direct focus to the specific challenges associated with this material and in particular, the reduction in the signal for serum albumin, antibodies and other major plasma proteins. In tissues, protein dynamic range is often not as exaggerated as in plasma. However, one tissue in particular, skeletal muscle, exhibits an asymmetry in the dynamic range of protein expression that is probably as extreme as blood plasma. This is due to two dominant fractions—the contractile apparatus and the sarcoplasmic fraction. The contractile apparatus (predominantly actin and myosin) is readily removed from skeletal muscle preparations as it is insoluble under low ionic strength conditions [28,29], but the residual component - the soluble protein fraction - also betrays a remarkable degree of specialization. One-dimensional or two-dimensional gel electrophoresis emphasizes the high level expression of a few tens of proteins, largely comprising the glycolytic enzymes responsible for fueling muscle contraction – see for example [30–35]. In fact, the soluble fraction from skeletal muscle might present as much of a challenge in terms of proteome dynamic range as does blood plasma.

It is thought that combinatorial equalization might not be compatible with absolute quantification, because of the finite capacity of the library for each protein, and because experiments are designed to suppress high abundance proteins, allowing measurement of low abundance components. Thus, high abundance proteins saturate the beads, and cannot be quantified. However, low abundance proteins might be accumulated on the library in sufficiently a linear fashion to permit sub-saturating quantification. That is only possible if the binding of the analyte is unaffected by the total protein exposure to the beads or the time of exposure. In this study, we applied the combinatorial equalization library to the soluble proteins of chicken skeletal muscle, a tissue that we have studied previously [31,35,36]. In particular, we set out to explore the outcome of an exaggerated over-saturation of the library to input protein pool. By the use of label-free [37–42] and QconCAT-based [35,43,44] analyses, we have uncovered behaviors, not previously seen in studies of extracellular fluids such as blood plasma, that may complicate the equalization of some types of cellular proteins.

**Methods**

**Sample preparation**

To isolate the soluble fraction of chicken skeletal muscle, supermarket purchased chicken breast (2 g) tissue was homogenized in 18 mL 20 mM sodium phosphate buffer, pH 7.0 containing protease inhibitors (Complete Protease Inhibitors, Roche, Lewes, UK). This was centrifuged at 15,000 g for 45 min at 4°C. The supernatant fraction, containing soluble protein, was then removed. The total protein concentration of the final preparation was measured using a Coomassie Plus Protein Assay (Pierce, Northumberland, UK).

**Equalization of proteins using Prospectrum-2 beads**

In preliminary sets of experiments, Prospectrum-2 library beads (a similar library is available as ProteoMiner™ Protein Enrichment Kits, BioRad, Hemel Hempstead, UK) were washed in 20 mg batches in 1 mL 50% (v/v) MeOH and mixed gently for 10 min. The beads were allowed to settle and the supernatant was removed and discarded. Methanol (50% (v/v)) was added to cover the surface of the beads that were then allowed to swell overnight at 4°C. Once swollen, 20 mg beads constituted 100 μL settled bed volume and this was transferred to a 1.5 mL Eppendorf tube. Beads were washed in 1 mL ddH₂O for 30 min prior to equilibration in 20 mM sodium phosphate buffer pH 7.0 for 30 min. After each wash, beads were allowed to settle for 5 min after which the supernatant was removed. Approximately 1 mL of protein mixtures containing 20 mg, 50 mg and 100 mg soluble protein from chicken skeletal muscle was added to the library in parallel experiments and mixed for 2 h. Unbound protein was collected as the supernatant fraction after the beads had been allowed to settle. The beads were subsequently washed five times in 20 mM phosphate buffer and supernatant fractions were removed and collected.

In a second series, designed to test the effect of extreme loading of beads, Prospectrum-2 beads, washed as described previously, were exposed to increasing amounts of soluble proteins from chicken skeletal muscle; chicken pectoralis (30 g) tissue was homogenized in 30 mL 20 mM sodium phosphate buffer, pH 7.0 containing protease inhibitors (Complete Protease Inhibitors, Roche, Lewes, UK). This was centrifuged at 15,000 g for 45 minutes at 4°C. The supernatant fraction, containing soluble protein was then removed. The total protein concentration of the final preparation was measured using a Coomassie Plus Protein Assay (Pierce, Northumberland, UK).

Increasing amounts of protein (20, 50, 100, 250, 500 and 1000 mg) were incubated with the library (in six parallel experiments) for 2 h. For increased protein loading (500 and 1000 mg), beads were incubated for an hour with less material (50% for 500 mg, 25% for 1000 mg), the unbound fraction was removed and the library was incubated with the same volume again of soluble protein; this was repeated four times for total exposure to the beads of 500 mg or 1000 mg protein. The beads were subsequently washed five times in 20 mM phosphate buffer and supernatant fractions were removed and collected. Beads were re-suspended in 200 μL phosphate buffer prior to analysis by 1D SDS-PAGE and mass spectrometry. Starting material, unbound protein, wash fractions and beads containing bound protein were analyzed using 1D SDS-PAGE. Approximately 10 μg protein from each fraction was loaded onto a 12.5% (w/v) acrylamide gel which was run at 200 V for 45 min.

In a final series, designed to test the effect on ionic strength on loading extent and complexity, increasing amounts of proteins (5, 10 and 25 mg) were incubated with 10 μL beads in one of two buffers; a low ionic strength buffer of 20 mM sodium phosphate buffer, pH 7.5 (ionic strength 50 mM) or a high ionic strength buffer of 20 mM sodium phosphate buffer, pH 7.5 containing 150 mM NaCl (ionic strength 200 mM). After incubation for 120 min at room temperature, the beads were washed five times in...
the binding buffer and the protein that remained attached to the beads was digested as described below for label-free quantification.

**Label free quantification**

For label-free quantification, samples (500 ng, 1 µL) from the protein loading study were resolved using a nano-ACQUITY system (Waters Corporation, Milford, MA, USA), equipped with a Symmetry C18 5 µm, 5 mm x 300 µm precolumn and an Atlantis C18 3 µm, 15 cm x 75 µm analytical RP column (Waters Corporation). The samples, 1 µL partial loop injection, were loaded with 0.1% formic acid solution at 4 µL/min for 3 min; mobile phase A (0.1% formic acid) and mobile phase B (0.1% formic acid in acetonitrile). After desalting and pre-concentration, the peptides were eluted from the precolumn to the analytical column and separated with a gradient of 3–40% mobile phase B over 90 min at a flow rate of 300 nL/min, followed by 10 min at 90% B. The column temperature was maintained at 35°C. Mass accuracy was maintained by use of a lock mass ([Glu1]-Fibrinopeptide B) that was delivered at 250 nL/min at a concentration of 100 fmol/µL to the reference spray.

Mass spectrometric analysis of tryptic peptides was performed using a Q-TOF Premier mass spectrometer (Waters Corporation). For all measurements, the mass spectrometer was operated in V-mode with a typical resolution of at least 10,000 FWHM. The TOF analyzer of the mass spectrometer was externally calibrated with a NaI mixture from m/z 50 to 1990. The data were post-acquisition lock mass corrected using the doubly charged monoisotopic ion of [Glu1]-fibrinopeptide B (m/z 785.8426). The reference spray was sampled at 60 s intervals. Accurate mass LC-MS data were collected in an alternating, low energy, and elevated-energy mode of acquisition (LC-MS5). The spectral acquisition time in each mode was 1.5 s with a 0.1 s interscan delay. In low energy MS mode, data were collected at constant collision energy of 4 eV. In elevated-energy MS mode, the collision energy was ramped from 15 to 40 eV during each 1.5 s integration. One cycle of low and elevated-energy data was acquired every 3.2 s. LC-MS data were processed and searched using ProteinLynx Global Server (PLGS) software, version 2.4. Protein identifications were obtained using the PLGS search engine, using an IPI *Gallus gallus* database. Identification required three fragment ion matches per peptide, seven fragment ion matches per protein and a minimum of one peptide match per protein. The digestion reagent was trypsin, one missed cleavage was permitted, the fixed modification was carbamidomethylation at cysteine residues and the variable modification was oxidation of methionine residues. For label-free quantification, yeast alcohol dehydrogenase (ADH, accession P00300, 50 fmol injected on column) was used as a standard. The three most intense peptides were then used for quantification relative to the ADH standard [42]. Each sample was run in triplicate, and the three triplicate LC MSE analyses were treated as independent data sets for hierarchical cluster analysis; in all instances, the triplicate data sets were clustered most closely, as would be expected. The individual analyses are retained in the presentation of the data. A decoy version of the database is generated ‘on the fly’ with every search conducted to infer peptide and protein level false positive identification rates. The allowed protein FDR was initially set at 4%, which typically accumulates to a peptide FDR<1%. However, none of the initial decoy identifications passed these filtering criteria; hence, the protein FDR and subsequently the peptide FDR of identification in the reported list of identifications are both close to zero.

**Quantification of equalized proteins using QconCAT**

Starting material and beads containing equalized protein were diluted 1:10 with ammonium bicarbonate (50 mM), to which [13C6]Arg, [15N2]lys-labelled QconCAT protein [35] was added (150 pmol to the starting material, 36 pmol to bead preparations containing equalized proteins). Protein was digested with trypsin at a ratio of protein: enzyme of 20:1 with incubation at 37°C for 24 h. To ensure complete digestion, digested protein was analyzed by 1D SDS-PAGE to confirm the absence of intact proteins.

**LC-MS**

For preliminary quantification, peptide mixtures were analyzed by LC-ESI-QTOF MS using an EASY-nLC (Proxeon, Odense, Denmark) nanoflow system coupled to a QTOF micro (Waters Corporation, Manchester, UK). Nanoflow HPLC at 200 nL/min was used to resolve peptides (in 0.1% v/v formic acid) over a 60 min acetonitrile gradient (0–100%). Peptides were acquired over the mass range 400–2000 m/z with the capillary voltage set at 1900 V, collision energy 10 V and sample cone at 55 V for the entire 60 min gradient. Q-peptides from proteins quantified as significantly enriched by the equalization process were confirmed by MS/MS (collision energy 30 V) and de novo sequencing. For later studies, peptides were analyzed by LC-ESI-LTQ MS/MS using an Ultimate 3000 HPLC system ( Dionex, Camberley, UK) coupled to a LTQ (Thermo Finnigan, Astmoor, UK). Nanoflow HPLC at 300 nL/min was used to resolve peptides (in 0.1% formic acid) over a 60 min acetonitrile gradient (0–100%). Peptides were acquired over the mass range 400–1500 m/z with the capillary voltage set at 50 V, spray voltage at 1.8 kV. Extracted ion chromatograms for heavy labeled Q-peptides were used for comparison of MS signal intensity of analyte and standard peptides.

**LC-MS/MS**

For protein identification, 1D gel separations of starting material and beads containing equalized proteins were divided into 22 slices, each of which was de-stained using 50:50 acetonitrile:50 mM ammonium bicarbonate, dehydrated with acetonitrile and digested overnight in-gel with trypsin. Resulting peptide solutions were analyzed by LC-ESI-LTQ MS/MS as described previously. MS/MS spectra were converted to dta files using BioWorks™ browser rev.3.3.1 (Thermo; MW range 400–3500, absolute threshold 10, precursor ion tolerance 2.5AMU, group scan 10, count 1, minimum ion count 1) and a Mascot generic format file was generated for all dta files. Data were searched against all Uniprot entries for *Gallus gallus* (database prepared on 30th April 2009, 10973 entries) using MASCOT (in house) from which only confident identifications (MOWSE score >34, p<0.05) were accepted. Search parameters consisted of trypsin as the proteolytic enzyme, one missed cleavage, no fixed modifications, oxidation of methionine as a variable modification, mass tolerance of the precursor ion set at 250 ppm with a tolerance of 0.6 Da for the fragment ions.

**Results and Discussion**

The soluble fraction of mature (approximately 30 d after hatch) chicken skeletal muscle comprises relatively few, high abundance proteins that are pronounced on 1-D SDS-PAGE but confirmed by 2-D GE. These are typically glycolytic enzymes [30,32,33,35,36] that dominate many proteome analyses, but which are particularly evident in muscle that has a high proportion of fast twitch 'white' fibres, such as the chicken pectoralis muscle (Figure 1a). The wash fractions were analyzed in the same way
Proteome Equalization

a) Proteins and Protein Loading

- GP (97kDa)
- PK (58kDa)
- B-eno (47kDa)
- Aldolase (38kDa)
- CK (43kDa)
- GAPDH (35kDa)
- TPI (27kDa)
- AK (22kDa)

SM U 25mg 50mg 100mg

Protein loading

b) Protein Concentration

- Starting material
- Beads + 50mg protein
- Beads + 100mg protein

Protein (nmol/g tissue)

TPI, MBC, β-actin, LDHB, TMA, E myo, GP, TMB, MHC, GAPDH, LDHA, α-eno, API, PK, CK

Protein Concentration

TPI, MBC, β-actin, LDHB, TMA, E myo, GP, TMB, MHC, GAPDH, LDHA, α-eno, API, PK, CK
to confirm that all unbound protein had been washed away from the beads (results not shown). The pattern of proteins on the gel changed considerably as proteins were loaded onto the beads. The largest differences were between the protein composition of the starting material and the bead contents at the lowest level of loading (25 mg). Several of the strongest bands on the gel in the starting material were absent from the loaded beads and new protein bands had appeared. As the loading doubled or quadrupled to 50 mg or 100 mg, the banding pattern continued to change, although the differences between the 50 mg and 100 mg loading were less pronounced. However, several major proteins were still evident on SDS-PAGE. To assess the change in protein composition of bead-bound proteins, 1D gels were sliced into segments and each slide was analyzed by LC-MS/MS (Figure S1, Table S1).

The extent of equalization was assessed by absolute quantification using stable isotope labeled internal standards. We have designed a QconCAT for the quantification of 20 high abundance soluble proteins in chicken skeletal muscle [35,43] to absolutely quantify the change in protein expression of these proteins during growth from 1–30 d. QconCAT was added to starting material or beads coated with equalized protein, prior to digestion with trypsin. When using surrogate peptides for absolute quantification it is vital that complete digestion is achieved and the digested protein samples were analyzed by 1D SDS-PAGE after digestion to ensure the complete removal of intact proteins (results not shown). For absolute quantification, peptides were analyzed by LC-ESI-MS using relative signal intensity of analyte (light) and internal standard (heavy) peaks. The identity of the peptides used for absolute quantification was confirmed by MS/MS and de novo sequencing (results not shown). This was expressed as nmol/g tissue before and after equalization, which after equalization lacks biological meaning but demonstrates in absolute terms the degree of protein equalization (Figure 1b,c) and permits comparison with previously published data. The extent to which the most abundant proteins, for example glyceraldehyde 3-phosphate dehydrogenase and $\beta$-enolase have been diluted during equalization is striking and other proteins, previously undetectable, were enriched (for example, tropomyosin A and actin polymerization inhibitor). This is exactly the behavior expected through use of the equalization beads.

To obtain a more comprehensive assessment of equalization, a fixed quantity of Prospectrum beads (20 mg dry weight) was incubated with soluble proteins at a total load of 25 mg, 50 mg, 100 mg, 250 mg, 500 mg and 1000 mg total soluble protein and each was incubated for 2 h. The beads were washed to remove unbound protein and starting material, beads and unbound protein were analyzed by label-free proteomics, using data independent MS$^3$ LC-MS acquisition, and combining the intensity of the top three precursor ions, giving protein identification data coupled with label-free quantification [37,41,42]. MS$^3$ based analysis is a data-independent approach that makes use of alternating low energy and high energy scans, conducted at high repetition rates. The data from the high-energy scan therefore contain fragment ions from all peptides present in the chromatographic flow. Post-acquisition, the precursor/product relationships are resolved by coincidence of retention time and by resolution of combinatorial possibilities of precursor/product relationships [39,40]. Moreover, the intensities of the three most intense peptide ions can be summed and yield, when compared to an internal standard digest (in these experiments, yeast alcohol dehydrogenase, accession P00330, 50 fmol applied to column), a reliable measure of the absolute abundance of each protein detectable in the mixture. Each loading was compared with the starting material, and was analyzed qualitatively (total proteins identified) and quantitatively (fmol each protein on column). Analyses were conducted in triplicate.

The un-equalised soluble protein fraction was first analysed by label-free proteomics using data independent data acquisition (MS$^3$) (Figure 2a). For the starting material of soluble muscle protein, a total protein load of $\sim$350 pmol protein yielded a very limited set of positive identifications, consistent with the marked bias in protein content in this tissue and extending over 2.5 orders of magnitude in dynamic range. As a fixed quantity of beads was exposed to increasing amounts of protein, the distribution of protein abundances became more shallow, and extended over a higher number of bound proteins. The equalization is evident from the distribution of the top 100 most abundant proteins in starting material (actually, only 35 proteins) and the highest bead loading – the markedly shallow profile is highly evident when the data are plotted on a linear scale (Figure 2b). This is concordant with the observation of increasing quantities of bead-bound protein, from 200 fmol on column at 20 mg protein load to over 500 fmol on column at the highest protein loads (Figure 2c). In the starting material, about 35 proteins could be confidently identified, a common experience in crude analyses of total skeletal muscle soluble protein. As the protein load increased, so the number of identifications followed suit, but reached a plateau at about 140 proteins from 250 mg loading to 10000 mg loading (Figure 2d; this is mirrored by the total protein calculated on column for the three replicate analyses of increased protein loading, panel c). The identified proteins are a little higher in number than the equivalent numbers obtained by LC-MS/MS analysis of individual gel slices but the limitations of both sets of identifications is consistent with restricted complexity and dynamic range in each sample.

The proteins that were identified changed throughout the loading range, consistent with the changes observed in the 1D SDS-PAGE patterns. A total of 360 proteins were identified using LC-MS$^3$ and 210 using GeLC-MS/MS (Table S2). To ensure the veracity of the label-free analysis in defining protein responses,
profiles from three proteins demonstrating different loading behaviors were analyzed by QconCAT and label-free quantification (Figure 3). The behavior of the three proteins was almost identical, irrespective of the quantification method used, lending credence to the comparative analyses. Creatine kinase, present at high levels in the starting material, was bound at very low levels to the beads; glycogen phosphorylase exhibited a biphasic response, initially binding at high levels but then at lower levels when the protein load was increased. Finally, actin continued to bind at very high levels as the beads were loaded with more and more input protein. The behavior of the last two proteins was unexpected, as there was no reason a priori why a protein should bind to fewer beads when the protein load was enhanced. Moreover, the failure of actin to saturate was also unexpected. Both of these behaviors implied that protein:protein interactions were taking place on the beads.

The comparison of QconCAT and label-free analysis validated the quantitative information from label-free analysis and the full data set for these proteins permitted a global exploration of the protein equalization of skeletal muscle proteins. When the entire data set, defining the abundance profiles for over 350 proteins and over seven loading levels was analyzed, hierarchical analysis sorted the different protein loadings according to the increasing mass (and concentration) of protein applied to the beads, and at all loading levels, correctly clustered the triplicate analyses together. In terms of the behavior of individual proteins, profile grouping by K-means clustering revealed four overall types of behavior. First, a small group of proteins (group C, Figure 4) were typified by very high concentrations in the starting material, but declined rapidly to practically undetectable levels as the beads were challenged with increasing quantities of muscle protein. These proteins, glyceraldehyde 3 phosphate dehydrogenase, enolase B, pyruvate kinase, lactate dehydrogenase A, aldolase A, phosphoglycerate kinase, creatine kinase, phosphoglucomutase, phosphoglycerate mutase, triose phosphate isomerase and adenylyl kinase are the same proteins that are highly abundant in the 1D SDS-PAGE gel of the soluble proteins of skeletal muscle (Figure 1). Typically, the quantities of these proteins bound to the beads declines to about 1% of that in the starting material. This is the behavior that would be expected for combinatorial equalization, since the few beads capable of binding each of these proteins would rapidly become saturated when exposed to such high concentrations of cognate proteins. Two further classes (A and B, Figure 4) demonstrated broadly the same behavior – the progressive appearance in analysis at heavy loading of protein onto the beads. Group B generally accumulated more readily than group A. Again, this would be consistent with the expected behavior of the beads.

However, a final class of proteins (Group D, Figure 4) was notable for unexpected and atypical behavior. This group of proteins was virtually undetectable in the starting material but accumulated on the beads to levels far higher than would be anticipated from the behavior of the majority of the other proteins. The proteins that exhibited this over-accumulation included heat shock proteins hsp90 and hsp70, α-actinin-2, calpains, initiation
factors (E1α-1) and calmodulins, proteins that were also identified in 1D SDS-PAGE analysis of specific bands (Figure 5). Contrary to expectations, the SDS-PAGE gel did not show a uniform ‘smear’ of a complex, approximately equimolar, protein mixture. There was strong evidence for selective binding and accumulation of specific proteins, as evidenced by intensely staining bands on the gel even at the highest loadings. These bands were excised and identified by LC-MS/MS, and their identities are indicated on the gel. There was thus good consistency between the label-free quantification and the developing asymmetry on gel-based analyses.

The dynamics of bead loading should be expected to be complex. As a limited set of binding sites are specific for each protein, loading would reflect binding of each bead with a mixture of proteins, the composition of which would be dictated by \( k_{on} \) and \( k_{off} \) and the solution concentration of the protein in the pool. At equilibrium loading, it would be anticipated that each site was fully saturated with specific proteins, resulting in a near-equimolar representation of the proteins present in the starting material. The intensity of protein bands in the pattern seen at 100 mg loading may reflect the saturation of beads with specific proteins that were present in adequate concentrations – as such, the density of these bands might reflect the intensity seen at saturation. If the beads were exposed to additional starting material, each protein reaching saturation might achieve the same intensity – true equalization. A proteome sample should then become progressively more equalized as the beads are exposed to greater and greater amounts of a protein mixture. Although low loadings of the beads had achieved some degree of equalization, we explored whether the degree of equalization would improve if a large excess of starting...
material was loaded on to the beads. In principle, more and more individual proteins should saturate the beads at approximately the same capacity. Notable proteins that were present in the lanes at higher loading included heat shock proteins, actinin and some very high molecular weight structural muscle proteins such as titin and connectin. None of these proteins were evident from LC-MS/MS analyses of the starting material, and therefore had been concentrated considerably by the beads.

Two explanations are offered for this behavior. First, the hexapeptide library on the bead might contain sufficient hydrophobic peptides to permit binding of proteins such as heat shock proteins to multiple beads, leading to asymmetric equalization. However, this is not a particularly satisfactory explanation for the other proteins that demonstrate similar behavior. The second explanation is that the beads can act as scaffolds for the assembly of complex networks that can accumulate in successive layers on the bead surface. The heat shock proteins are able to bind a wide range of proteins with exposed hydrophobic surfaces, and their marked predominance in the heavily loaded beads suggests that this is the case. The marked enhancement of HSP90 and α-actinin imply the development of a large assembly of protein complexes consisting primarily of these proteins, implying a preferential association between these molecules.

The interaction between the hexapeptides in the library and the target proteins should be expected to vary from protein to protein. Electrostatic interactions would be anticipated to diminish at high ionic strength, whereas for salts such as NaCl used here, hydrophobic interactions should be strengthened. The pattern of protein binding could therefore be expected to depend on solution conditions. To test this, portions of the hexapeptide beads were incubated with skeletal muscle soluble proteins in two buffers, one at low ionic strength (50 mM) and one at high ionic strength (200 mM). Slightly more protein was bound to the beads at high ionic strength (Figure 6a), behavior that would be consistent with enhancement of hydrophobic interactions. At these low protein loadings, similar numbers of proteins were bound, with slightly more proteins being retained by the beads at the higher ionic strength. However, this overall description conceals the variation elicited by ionic strength. Of 222 proteins bound to the beads at 25 mg load, 107 were exclusively bound at high ionic strength, 41 proteins were bound at low ionic strength and 74 were bound at both ionic strength values. For the proteins bound at both solution conditions, the ratio of the quantity of protein bound at high:low ionic strength values varied considerably, from over 30:1 (lactate dehydrogenase A) to 0.03:1 (titin). It is clear that the solution conditions can have a profound influence on the protein binding profile, which opens up new opportunities for extending the scope and range of the equalization profiles. To illustrate the difference in behavior of individual proteins at different ionic strength values, representative proteins are included in Figure 6. Some proteins, notably the glycolytic enzymes such as glyceraldehyde-3-phosphate dehydrogenase or

Figure 5. Equalization achieved by extreme bead loading. Prospectrum II beads were exposed to very high levels (up to 1000 mg protein/20 mg beads) of chick skeletal muscle soluble proteins in a final volume of 1.0 mL of 20 mM sodium phosphate buffer, pH 7.0. After 2 h, the beads were washed extensively (8 successive washes in the same buffer) and samples were analyzed by 1D SDS-PAGE as described in the legend to Figure 1. Bands were excised, digested with trypsin and analyzed by LC-MS/MS and selected identified proteins are highlighted on the gel. doi:10.1371/journal.pone.0028902.g005
glycogen phosphorylase are bound to a greater extent in high ionic strength buffers. Others, such as the heat shock proteins and titin, bind much more extensively at low ionic strength values. Finally, proteins such as alpha actinin are bound strongly, irrespective of the solution conditions. Careful selection of solution conditions might therefore not only diminish protein assemblies but also be used to extend the range of proteins that are accessible. This might usefully be extended in an orthogonal complementarity to adjustment of pH, as has recently been successfully demonstrated [27,45].

Most studies using bead-based proteome equalization have focused on biological fluids, and in particular plasma although recently, these equalization analyses have been extended to some rather more exotic proteomes [46–48]. Although it has a high dynamic range of protein expression, plasma does not contain high concentrations of heat shock proteins, and, almost by virtue of their function in plasma, true plasma proteins might not be anticipated to have exposed hydrophobic surfaces, although serum albumin is known to have exposed hydrophobic regions. However, any tissue sample would be anticipated to be replete with heat shock proteins, and the behavior that is observed here with the soluble proteins of skeletal muscle might be replicated with other tissue samples. For cellular material or tissues, the quest for the ‘democratic proteome’ [18] may well be thwarted by the inability of some proteins to behave as anticipated, democratically, and indeed, subvert the behavior of other proteins. If the hypothesis of a heat shock protein framework turns out to be proven in other tissues, then equalization of tissue proteomes may remain elusive, unless bead-loading conditions can be elucidated to minimize or eliminate the effect.

Supporting Information

Figure S1 Protein identification following equalization of chicken skeletal muscle proteins. For protein identification, 1D gel separations of starting material and beads containing equalized proteins were divided into 22 slices, each of which was de-stained and digested overnight in-gel with trypsin. Resulting peptide solutions were analyzed by LC-ESI-LTQ MSMS and MSMS data were searched against all Uniprot entries for *Gallus gallus* (database prepared on 30th April 2009, 10973 entries) using MASCOT from which only confident identifications (MOWSE score>50, p<0.05) were accepted, for details see Table S1. (PDF)
Table S1  Protein identifications derived from in-gel analyses.

Table S2  Protein identification and label-free quantification of bead-bound proteins.

Acknowledgments

We are grateful to Dave Hammond, DeAnna Long and Julia Lathrop (Peptides international) for their advice, also to Dr Gary Evans (Genus plc) for his interest in this work. Part of this work was conducted whilst RJB was engaged in sabbatical studies at Waters plc, and the instrument training and technical support of colleagues at this site is gratefully acknowledged.

Author Contributions

Conceived and designed the experiments: RJB TM JIL JR. Performed the experiments: JR RJB CH YW. Analyzed the data: RJB JPCV JR YW. Contributed reagents/materials/analysis tools: JR JPCV YW. Wrote the paper: RJB JR JIL JPCV.

References

1. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1: 845–867.
2. Anderson NL, Polanski M, Peeper R, Gatlin T, Tirunalmal RS, et al. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. Mol Cell Proteomics 3: 311–326.
3. Brunsvig P, Holman SW, Gaskell SJ, Grant CM, Harman VM, et al. (2011) Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. Proteomics 11: 2957–2970.
4. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, et al. (2003) Global analysis of protein expression in yeast. Nature 425: 737–741.
5. Zhang H, Liu AY, Loriaus P, Wolfscheid B, Zhou Y, et al. (2007) Mass spectrometric detection of tissue proteins in plasma. Mol Cell Proteomics 6: 64–71.
6. Gronwall C, Sjoberg A, Ramstrom M, Houdin-Guthenich I, Hober S, et al. (2007) Affibody-mediated transferin depletion for proteomics applications. Biotechnol J 2: 1309–1308.
7. Ramstrom M, Zuberovic A, Gronwall C, Hanrieder J, Hanrieder K, et al. (2009) Development of affinity columns for the removal of high-abundance proteins in preclinical trials. Biotechnol Appl Biochem 52: 159–166.
8. Pavanin RR, Beaudoin E, Bucham M, Beliveau R (2007) Proteomic analysis of human plasma proteins by two-dimensional gel electrophoresis and by antibody arrays following depletion of high-abundance proteins. Cell Biochem Biophys 49: 102–193.
9. Liu T, Qian WJ, Mottaz HM, Grisenko MA, Norbeck AD, et al. (2006) Evaluation of multiprotein immunoenrichment affection for plasma proteomics and candidate biomarker discovery using mass spectrometry. Mol Cell Proteomics 5: 2167–2171.
10. Plavina T, Wakoahall E, Hancock WS, Hicapnie M (2007) Combination of abundant protein depletion and multiloci affinity chromatography (M-LAC) for plasma biomarker discovery. J Proteome Res 6: 662–671.
11. Zolotarova N, Martosella J, Nicola G, Bailey J, Boyes BE, et al. (2005) Differences among techniques for high-abundant protein depletion. Proteomics 5: 3304–3313.
12. Candiano G, Dimuccio V, Bruchi M, Santucci L, Gusmania R, et al. (2009) Combinatorial peptide ligand libraries for urine proteome analysis: investigation of different elution systems. Electrophoresis 30: 2405–2411.
13. Castagna A, Cecconi D, Semich L, Rapsibelli J, Guerrier L, et al. (2005) Exploring the hidden human urinary proteome via ligand library beads. J Proteome Res 4: 1917–1930.
14. Guerrier L, Righetti PG, Boschetti E (2008) Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteomes by means of hexapeptide ligand library. Nat Protoc 3: 885–890.
15. Guerrier L, Thulasiramam V, Castagna A, Fortis F, Lin S, et al. (2006) Reducing protein concentration range of biological samples using solid-phase ligand libraries. J Chromatogr B Analytical Technol Biomed Life Sci 833: 33–40.
16. Righetti PG, Boschetti E, Kaschak AV, Fasoli E (2010) The proteome buccaneers: how to unearh your treasure chest via combinatorial peptide ligand libraries. Expert Rev Proteomics 7: 373–385.
17. Righetti PG, Boschetti E, Lomass M, Citterio A (2006) Protein Equalizer Technology : the quest for a “demonic proteome”. Proteomics 6: 3980–3992.
18. Righetti PG, Boschetti E, Zanella A, Fasoli E, Citterio A (2010) Phciking, pillaging and plundering proteomes with combinatorial peptide ligand libraries. J Chromatogr A 1217: 983–990.
19. Righetti PG, Castagna A, Antonelli P, Boschetti E (2005) Prefractionation techniques in proteome analysis: the mining tools of the third millennium. Electrophoresis 26: 297–319.
20. Semich L, Salek M, Lomass M, Boschetti E, Righetti PG, et al. (2007) Proteinomic analysis of human blood serum using peptide library beads. J Proteome Res 6: 4055–4062.
21. Righetti PG, Boschetti E (2007) Sherlock Holmes and the proteome—a detective story. FEBS J 274: 957–965.
45. Fasoli E, Farinazzo A, Sun CJ, Kravchuk AV, Guerrier L, et al. (2010) Interaction among proteins and peptide libraries in proteome analysis: pH involvement for a larger capture of species. J Proteomics 73: 733-742.
46. D’Amato A, Fasoli E, Kravchuk AV, Righetti PG (2011) Mehercules, ad hoc Bacchus! The Debate on Wine Proteomics Continues. J Proteome Res 10: 3789-3801.
47. Cereda A, Kravchuk AV, D’Amato A, Bachi A, Righetti PG (2010) Proteomics of wine additives: mining for the invisible via combinatorial peptide ligand libraries. J Proteomics 73: 1732-1739.
48. D’Amato A, Bachi A, Fasoli E, Boschetti E, Peltre G, et al. (2009) In-depth exploration of cow’s whey proteome via combinatorial peptide ligand libraries. J Proteome Res 8: 3925–3936.