

REVIEW

Open Access

Quality assessment and optimization of purified protein samples: why and how?

Bertrand Raynal^{1,2*}, Pascal Lenormand^{1,2}, Bruno Baron^{1,2}, Sylviane Hoos^{1,2} and Patrick England^{1,2*}

Abstract

Purified protein quality control is the final and critical check-point of any protein production process. Unfortunately, it is too often overlooked and performed hastily, resulting in irreproducible and misleading observations in downstream applications. In this review, we aim at proposing a simple-to-follow workflow based on an ensemble of widely available physico-chemical technologies, to assess sequentially the essential properties of any protein sample: purity and integrity, homogeneity and activity. Approaches are then suggested to optimize the homogeneity, time-stability and storage conditions of purified protein preparations, as well as methods to rapidly evaluate their reproducibility and lot-to-lot consistency.

Keywords: Recombinant protein, Therapeutic, Diagnosis, Structural biology, Electrophoresis, Mass spectrometry, UV/visible spectroscopy, Light scattering, Size-exclusion chromatography, Surface plasmon resonance, Formulation, Comparability

Introduction

In recent years, purified proteins have more and more frequently been used for diagnostic and therapeutic applications [1-3]. Purified proteins are also widely used as reagents for downstream in depth biophysical and structural characterization studies: these are sample- and time-consuming, generally requiring long set-up phases and sometimes depending on (limited) accessibility to large instrumentation such as synchrotrons.

Unfortunately, scientists (especially in the academic environment) frequently want to rush to the final application, considering biochemical analysis of proteins as either trivial or a superfluous bother. Very often, the implications of such a regretful attitude are irreproducible, dubious and misleading results, and unfortunately sometimes lead to failure at more or less advanced stages (including clinical trials [4]), with potentially severe consequences. This is even more the case nowadays, when recombinant production of challenging proteins such as integral membrane proteins or heavily modified (glycosylated, ...) proteins is being attempted on an ever more widespread scale.

The correct interpretation of many biophysical/structural characterization experiments relies on the assumption that:

- 1) the protein samples are pure and homogeneous.
- 2) their concentration is assessed precisely.
- 3) all of the protein is solubilized and in a natively active state.

Our experience as a core facility dealing with several dozens of different projects every year is that quality control considerations are much too often overlooked or taken for granted by facility users and the scientific community at large. However, those who assess and optimize carefully the quality of their protein preparations significantly increase their chances of success in subsequent experiments.

Purified protein quality control has already been the object of several general reviews [5-7]. Attempts have also been made to define a set of "minimal quality criteria" that should be fulfilled by any purified recombinant

* Correspondence: bertrand.raynal@pasteur.fr; patrick.england@pasteur.fr

¹Institut Pasteur, Biophysics of Macromolecules and their Interactions, 25 rue du Docteur Roux, 75724 Paris Cedex15, France

²CNRS-UMR3528, Institut Pasteur, Departement of Structural Biology and Chemistry, Paris, France

protein prior to publication, especially among the “Minimal Information for Protein Functionality Evaluation” (MIPFE) consortium [8-10]. In this review, we wish to go one step further and provide a concise overview of a sequence of simple-to-follow physico-chemical approaches that should be accessible to the vast majority of investigators. Most of the methodologies that are proposed can be found in classical biochemistry or structural biology laboratories, and in the majority of institutional protein science core facilities. Many of the methods and techniques mentioned here are well known, maybe too well, but clearly need to be reappraised in university curricula and laboratory practice: indeed knowledge about them is generally (and inappropriately) regarded as obvious, but very often it is in reality very sketchy, sometimes unfortunately resulting in gross blunders. Hopefully, this review will help providing more robustness to the production of efficient and reliable protein samples within a large scientific community.

Protein quality control methodological work-flow

Initial Sample assessment

Purity and integrity

Electrophoresis Prior to any downstream experiment, purity and integrity are the very first qualities that need to be assessed for any protein sample (Figure 1B). This is routinely achieved by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). This technique, associated with Coomassie blue staining, can detect bands containing as little as 100 ng of protein in a simple and relatively rapid manner (just a few hours) [11]. After reduction and denaturation by SDS, proteins migrate in the gel according to their molecular mass, allowing to detect potential contaminants, proteolysis events, etc. However, many low amount impurities and degradation products can go unnoticed, especially in low concentration samples or during optimization phases in which minute aliquots are analysed.

Two higher sensitivity colorimetric staining methods can be used either directly after electrophoresis or coupled to Coomassie blue staining: zinc-reverse staining [12] and silver staining [13]. These can detect as low as 10 ng and 1 ng protein bands respectively. Zinc-reverse staining (also known as negative staining) uses imidazole and zinc salts for protein detection in electrophoresis gels [12]. It is based on the precipitation of zinc imidazole in the gel, except in the zones where proteins are located. When zinc-reverse staining is applied on a Coomassie blue stained gel, previously undetected bands can be spotted [14]. This technique is rapid, simple, cheap and reproducible, and is compatible with mass spectrometry (MS) [15]. On the other hand, silver staining is based on the binding of silver ions to the proteins followed by reduction to free silver, sensitization and enhancement [13]. If used as a second staining, it is essential to fix the proteins in the gel with

acidic alcohol prior to initial Coomassie blue staining [16]. Two drawbacks of this technique are that proteins are differentially sensitive to silver staining and the process may irreversibly modify them preventing further analysis. In particular glutaraldehyde, which is generally used during the sensitization step, may interfere with protein analysis by MS due to the introduction of covalent cross-links [17]. To circumvent this problem, a glutaraldehyde-free modified silver-staining protocol has been developed, which is compatible with both matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization-MS [17].

Several fluorescent dyes such as Nile red, ruthenium(II) tris(bathophenanthroline disulfonate) (RuBPS), SyPro and Epicocconone, can also be used to reveal a few ng of proteins in gels [18-20]. CyDyes can even reveal amounts of protein lower than one nanogram but have the inconvenience of requiring to be incorporated before gel electrophoresis [20]. Apart from Nile red, these staining methods are compatible with subsequent MS analysis. However, their major disadvantage is that they require a fluorescence imager for visualization and that they are significantly more expensive than classical colorimetric dyes.

Different alternatives (or additions) to SDS-PAGE exist to further separate and distinguish the protein of interest from closely related undesired subproducts or contaminants. One of them is isoelectric focusing (IEF), which separates non-denatured proteins based on their isoelectric point, most often on gel strips. This allows to resolve proteins of very similar mass, notably unmodified and small molecular mass post-translationally modified (e.g. phosphorylated) variants of a same protein. IEF is often used upstream of SDS-PAGE in so-called 2D gel electrophoresis [21].

Capillary electrophoresis (CE) is another useful alternative, with the advantage of superior separation efficiency, small sample consumption, short analysis time and automatability. CE separates proteins, with or without prior denaturation, in slab gels or microfluidic channels, according to a variety of properties, including their molecular mass (SDS-CGE), their isoelectric point (CIEF) or their electrophoretic mobility (CZE) [22]. Interestingly, CE can readily be coupled on line with MS [23].

UV-visible spectroscopy UV-visible spectroscopy is most often used for protein concentration measurements (see Total protein concentration determination section). However, it is also a very convenient tool for the detection of non-protein contaminants, as long as the protein of interest contains aromatic residues and the absorbance is monitored over a large range (at least 240 – 350 nm). In particular, undesired nucleic acid contaminants can be spotted as bumps at 260 nm, resulting in a high 260/280 nm absorbance ratio (which should be close to 0.57 for a non-contaminated protein sample [24]). On the other hand, reducing agents (especially DTT) alter the

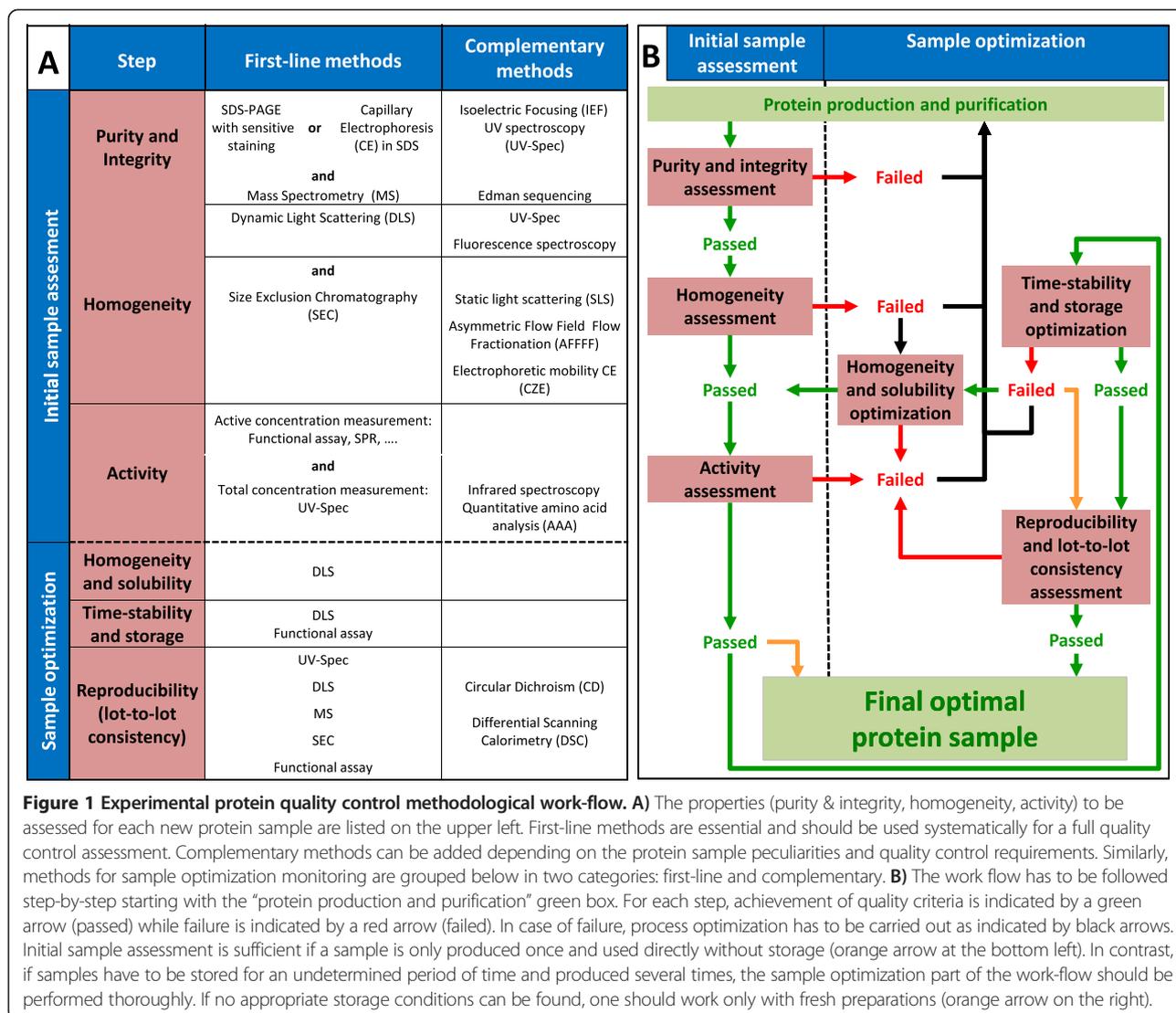


Figure 1 Experimental protein quality control methodological work-flow. **A**) The properties (purity & integrity, homogeneity, activity) to be assessed for each new protein sample are listed on the upper left. First-line methods are essential and should be used systematically for a full quality control assessment. Complementary methods can be added depending on the protein sample peculiarities and quality control requirements. Similarly, methods for sample optimization monitoring are grouped below in two categories: first-line and complementary. **B**) The work flow has to be followed step-by-step starting with the “protein production and purification” green box. For each step, achievement of quality criteria is indicated by a green arrow (passed) while failure is indicated by a red arrow (failed). In case of failure, process optimization has to be carried out as indicated by black arrows. Initial sample assessment is sufficient if a sample is only produced once and used directly without storage (orange arrow at the bottom left). In contrast, if samples have to be stored for an undetermined period of time and produced several times, the sample optimization part of the work-flow should be performed thoroughly. If no appropriate storage conditions can be found, one should work only with fresh preparations (orange arrow on the right).

symmetry of the 280 nm absorbance peak by increasing the absorbance at 250 nm and below [25,26].

Mass spectrometry It is essential to verify the integrity of the protein of interest beyond SDS-PAGE, especially when setting-up a new production/purification protocol, as low level proteolysis events (affecting just a few amino acids) and undesired modifications may go unnoticed in electrophoresis. The method of choice for detailed analysis of protein primary structure is MS, as it can provide molecular mass with 0.01% accuracy for peptides or proteins with masses up to 500,000 Da using only a few picomoles of sample [27]. The presence of undesired proteolytic events and chemical alterations can be readily detected by comparing the difference between the observed and the expected mass of the protein. Furthermore MS can provide detailed information about the presence of desired post-translational modifications (phosphorylations, acetylations,

ubiquitinations, glycosylations, ...) [28]. Overall the convenience and precision of MS measurements is such that they should be considered as routine to ensure the integrity and overall state of modification of the peptide or protein of interest.

MS-based methods, such as MALDI in-source decay [29], are progressively replacing traditional protein sequencing by Edman degradation [30]. However, N-terminal Edman sequencing is still of relevance in several cases, for instance when one wishes to verify easily and specifically the N-terminal boundary of the protein of interest, or when highly accurate masses cannot be obtained by MS because of the size of the protein or the presence of certain post-translational modifications [31].

One may also wish to further characterize the degradation products or contaminants detected by electrophoresis, as determining their origin may give clues about how to avoid them from occurring. Proteins extracted from gel

bands can be digested and analysed by MS [32]. Identification can be achieved by peptide mass finger-printing, as the precise peptide pattern that results from the digestion of a protein by a sequence-specific protease (like trypsin) is unique for each protein and can be matched by protein-sequence database search [32]. Usually MALDI time-of-flight (TOF) spectrometers are used for this type of analysis because of their speed, mass accuracy and sensitivity. Typically, proteins detected by Coomassie blue or negative staining can be identified.

Homogeneity

Dynamic light scattering Once the purity and integrity of the protein sample has been assessed, one has to ensure it is homogeneous (Figure 1). Dynamic light scattering (DLS), because of its rapidity and low sample consumption, is a very convenient method to determine simultaneously the monodispersity of the species of interest and the presence of soluble high-order assemblies and aggregates [33]. DLS measures Brownian motion, which is related to the size of the particles. The velocity of the Brownian motion is defined by a translational diffusion coefficient that can be used to calculate the hydrodynamic radius, i.e. the radius of the sphere that would diffuse with the same rate as the molecule of interest. This is done by measuring, with an autocorrelator, the rate at which the intensity of the light scattered by the sample fluctuates. As a 3 nm radius particle scatters 1 million times less light than a 60 nm one, DLS is the method of choice to detect small quantities of aggregates in a sample [34]. A few percent of large aggregates may even swamp the scattered light coming from small particles. It is important to notice that large particles may also originate from poor buffer preparation (all protein purification and storage buffers should systematically be filtered prior to use). Autocorrelation functions can be mathematically resolved using a variety of algorithms, developed either by instrument manufacturers or academic researchers (for instance Sedfit [35]). However, the robustness of these mathematical solutions is fairly poor. Moreover, a precise quantification of each individual species is difficult and the resolution of DLS does not allow to resolve close quaternary structures (for instance monomers from dimers and small-order oligomers). Overall, DLS is such an easy and convenient technique that the danger of over-interpreting its quantitative results is high [34]. However, the technique is very well adapted for qualitative studies (which are the focus of this review) and can be performed over time and/or at different temperatures in order to test the stability of the protein preparation in different buffers (see Optimization of homogeneity and solubility section).

UV-visible and fluorescence spectroscopies Although less sensitive than DLS, UV-visible spectroscopy is also of use to detect the presence of large particles (with a hydrodynamic radius higher than 200 nm) in a protein preparation. This can be done by monitoring the absorbance signal above 320 nm, where aggregate-free protein samples are not supposed to absorb light, and the signal can be attributed exclusively to the scattering of light by large aggregates present in the sample. This simple measurement can quickly provide qualitative information about the sample of interest. If the UV visible signal is used for concentration measurement, the contribution of scattering to the overall absorbance can be deduced by tracing a log-log plot of absorbance versus wavelength in the 320–350 nm region. This can then be extrapolated to the rest of the spectrum [26,36].

One interesting alternative to UV-visible spectroscopy is fluorescence spectroscopy [37]. After excitation at 280 nm, the fluorescence emission signal is measured at 280 nm and 340 nm, corresponding respectively to light scattering and intrinsic protein fluorescence. The ratio of the intensities at 280 nm and 340 nm (I_{280}/I_{340}) is concentration independent and purely related to the degree of aggregation of the sample. This ratio, also called aggregation index (AI), should be close to zero for aggregate-free protein preparations and can attain high values (>1) when significant aggregation occurs.

Size-exclusion chromatography As already stressed above, DLS does not have the sufficient resolution to correctly assess whether a protein sample is heterogeneous in terms of oligomerisation. Analytical size exclusion chromatography (SEC) is currently the standard separation technique to quantify protein oligomers. SEC, which very often is also the last step of protein purification, separates molecules according to their hydrodynamic size, often defined by their Stokes or hydrodynamic radius [38], with larger sized molecular species (which are not necessarily larger molecular mass species) eluting before smaller ones. Recent developments of the technique have increased the rapidity of elution, through column parallelization and injection interlacing [39] and/or the use of the latest SEC columns with smaller pore size, allowing improved resolution with smaller bed volumes, reduced elution times (below 10 min) and low sample consumption (5 μg in 20 μl) [40–42]. This should encourage people to resort to SEC as a systematic approach to analyse sample heterogeneity. Aggregates, contaminants and potentially different molecular arrangements of the protein of interest can be readily separated and quantified, with classical online UV detection. One should however keep in mind the fact that the protein sample will be diluted during SEC by as much as a 10-fold factor, which might alter equilibria between oligomeric species.

Furthermore, however “inert” may be the gel filtration resins, some proteins do interact with them, rendering SEC impossible. Two column-free separation techniques may be used as alternatives: asymmetric flow-field flow fractionation (AFFFF), which is also well suited for large molecular assemblies that may be dissociated by SEC [42,43], and capillary electrophoresis with electrophoretic mobility separation (CZE) [22].

Static light scattering Contrary to a widespread belief, the molecular mass of the species eluted in each SEC peak cannot be obtained through column calibration approaches, in which protein standards are separated according to their hydrodynamic radius and not their molecular mass (the correlation between both parameters being far from linear, especially for non-globular and intrinsically disordered proteins). To obtain information about mass, it is necessary to resort to a static light scattering (SLS) detector [44], in combination with a UV or a refractive index (RI) detector. Of note, as in the case of DLS, SLS is also able to detect small amounts of aggregates with high sensitivity, as the light scattering signal is proportional to molecular mass [45]. In size exclusion chromatography with on-line static laser light scattering (SEC-SLS), experimentally determined molecular mass is independent of the elution volume of the protein. Both the total scattered light intensity (which depends on molecular mass and concentration) and the concentration of the protein (using the UV or RI detector) are measured and analysed to determine the molecular mass of the protein as it elutes from the chromatographic column. SEC-SLS is applicable and quite accurate over a broad range of molecular masses (from a few kDa to several MDa), as long as the column is able to resolve completely the different species present in the sample, allowing the area of each peak to be integrated. In order to improve the separation of peaks with respect to traditional SEC, one can resort to ultra-high performance liquid chromatography (UHPLC) systems, which have very recently been made amenable to SLS. As an alternative, AFFF can also be used in conjunction with SLS [42,43].

Activity

Active protein concentration determination Once the homogeneity of the protein of interest has been assessed, one has to ensure it is active and functional (Figure 1). An infinite variety of generic or protein-specific functional assays has been designed, relying principally on catalytic and binding properties. An attempt at listing such assays would go much beyond the scope of this review. Efficient assays allow to measure precisely the active concentration of the protein sample, and thus to determine (if the total protein concentration is known: see Total protein concentration determination section) the percentage of purified protein that is indeed functional. One should not overlook such active

protein concentration determinations, as it can unfortunately often be found that the proportion of purified protein which is indeed in a native active state is low. This can be due to misfolding issues, to the inability of the protein to reach its native structural state spontaneously or to interferences of sequence additions (such as tags or extra amino acids originating from cloning vectors). But in most cases, this is due to poor (and overlooked) micro-integrity and homogeneity of the purified protein (see Purity and integrity section).

Surface plasmon resonance (SPR) is a convenient technique to determine the active concentration of binding proteins. This is done by exploiting the properties of diffusion of molecules in continuous flow microfluidic devices [46,47]. The so-called “calibration-free concentration analysis” (CFCA) method, which has been implemented in a user-friendly format in different SPR instruments available commercially [48], allows to determine the concentration of protein able to recognize a specific ligand (or protein partner) tethered on a surface. For CFCA measurements, the ligand has to be immobilized at high densities, creating conditions in which the interaction rate of the protein is limited by its diffusion towards the surface (mass transport limitation), and becomes proportional to its active concentration [46,47].

Alternatively, if the protein of interest is tagged, one can resort to a “sandwich” SPR assay to determine directly what proportion of protein is active: a measurable amount of protein is first captured through its tag on a surface on which a tag-specific receptor is immobilized (NTA for His-tag, or an antibody for others) and then titrated by a saturating amount of specific ligand [49].

Total protein concentration determination Different methods are available to measure the total protein concentration in a sample, allowing to deduce the percentage of active protein (see Active protein concentration determination section). Bradford, bicinchoninic acid (BCA) and Lowry assays use standards for calibration, which can be a source of error as the composition of the protein of interest may not necessarily match that of the protein standards [26]. It is also possible to use UV-visible absorbance measurements to determine the total protein concentration as long as its extinction coefficient is reliably known or calculated [26,50]. The extinction coefficient at 280 nm is most frequently calculated from the amino acid composition [25], allowing to determine concentrations from UV absorbance at this wavelength (see [26,50] for protocols). However, one should always monitor wider absorbance spectra (at least from 240 to 350 nm), as these can provide much more information than concentration, as already detailed in the two sections referring to UV-visible spectroscopy above.

However, UV absorbance measurements are only usable for concentration determination if the sequence of

the protein of interest contains a known amount of tryptophans and tyrosines, the two principal light-absorbing amino acids. If this is not the case, an alternative is to use Fourier Transform Infrared Spectroscopy (FTIR) as initially suggested by Etzion et al. [51]. After subtracting the contribution of water between 1700 nm and 2300 nm, the analysis of the amide band I and II of the IR absorbance spectrum can be used to calculate protein concentration by determining the concentration of amine bonds. Recently, commercially available FTIR equipment has been developed (Direct Detect from Merck Millipore), applying this method to protein samples that are dried on a membrane. The only limitations of the equipment are the minimal and maximal concentrations that can be used (0.2 to 5 mg/ml) and the incompatibility of several amine-containing buffers (HEPES \geq 25 mM, Tris \geq 50 mM, ...) or additives (EDTA \geq 10 mM, ...). Another alternative is amino acid analysis (AAA) which is a very valuable technique both for protein identification and quantification [52]. Briefly, quantitative AAA involves hydrolyzing the peptide bonds to free individual amino acids, which are then separated, detected and quantified, using purified amino acids as standards (see [52] for protocol).

Nonetheless, UV-visible spectroscopy remains beyond any doubt the most widely spread, cost- and time-efficient technique for total protein concentration determination. To take full advantage of this technique even in the absence of tyrosine and tryptophan residues, one solution can be to use FTIR-based protein quantification and AAA measurements at first, to generate concentration calibration curves for the protein of interest in correlation with UV absorbance (at 280 nm or another wavelength). These calibration curves can then be used to determine the concentration of subsequent samples directly by UV absorbance spectroscopy.

Optimization, stability and reproducibility of protein samples

Identifying conditions in which a protein sample is "well-behaved" and meets all the required criteria described in Initial sample assessment section is generally not a trivial task. In this section, we aim at providing an overview of potential solutions to overcome difficulties that may arise along the quality control work-flow (Figure 1). We also discuss how to determine optimal conditions for the preservation of good quality samples, and how to ensure that the protein production/purification process that one has devised leads reproducibly to samples of equivalent high quality.

Optimization of purity and integrity

A variety of solutions are available to overcome issues of contamination of protein samples with impurities, degradation products or undesired chemically-modified

proteins [53]. These go from changing the purification protocols (modifying the washing and elution conditions from affinity chromatography columns, or adding purification steps such as ion-exchange chromatography) to more upstream changes such as the addition of different sets of protease inhibitors, the modification of the conditions of induction of protein expression, the choice of another cloning vector (with a different tag, or a tag placed at another position or at both ends), or even resorting to another expression host system.

Optimization of homogeneity and solubility

To remove protein aggregates, it is important to ensure that the last step of the purification process always is size-exclusion chromatography. A column should be chosen that allows elution of the protein of interest well away from the void volume, and thus total separation from large protein aggregates. People often need to concentrate their protein samples in order to attain concentrations high enough for their downstream applications: unfortunately, this process, which resorts to spin concentrators or precipitation/resolubilisation protocols, very frequently tends to induce aggregation. Therefore, one should be careful not to concentrate their sample more than strictly necessarily (avoiding overly high concentrations): this should either be done *before* the final size-exclusion chromatography step, or be followed by an analytical SEC or DLS on part of the concentrated sample to ensure that it has remained free from aggregates.

To minimize the formation of protein aggregates (and to improve solubility), a variety of changes can be made upstream to the production/purification protocol [54]. Adjustment of several parameters of the sample buffer composition (pH, salinity, presence of additives, co-factors or ligands, ...) can also dramatically increase homogeneity. People often rely for this on empirical rules that they have learnt with experience, as there is no clear correlation between the stability of a protein and its intrinsic properties (amino acid composition, isoelectric point, secondary structure elements, ...). Recent DLS instrumental developments, that allow to process a large number of samples in a 96, 384 or 1536 well plate format, have made buffer condition screening an easy task. Many groups have used DLS as a technique to improve the solubilisation conditions of their proteins, in particular before crystallization studies [55,56]. Buffer matrices for multi-parametric screening of pH, salinity, buffer nature, additives and co-factors can be generated by hand or using simple robotics [57]. Typically samples, at a concentration of 10 mg/ml for a 10 kDa protein or 1 mg/ml for a 100 kDa protein, are diluted 10 times in each test buffer with a consumption of only 2 μ l of sample per condition. The homogeneity of the sample and the presence of aggregates (and high-order

physiologically irrelevant oligomers) can be monitored in each condition, allowing to select the optimal buffer composition for protein homogeneity.

Optimization of protein sample stability and storage

Preservation of good quality protein samples over time is all important, as very often one will not consume all of a sample straight away. People most often rely on hearsay for the short-term or long-term storage of their precious protein samples. A very widely spread belief is that flash freezing (with or without cryoprotectants such as glycerol) is the best method for long-term retention of protein properties. However, this is far from being a general truth, especially because significant denaturation, aggregation and precipitation can occur upon freezing/thawing [58]. Proteins may become unstable and lose their biological activity through a variety of physical or chemical mechanisms, even at cold temperatures [59-61]. The best storage conditions are very much protein-dependent, and may vary from unfrozen aqueous solutions to salted precipitates or freeze-dried solids [59-61].

A practical way to approach this issue is to start by monitoring the time stability of one's protein sample at a few relevant temperatures (e.g. 4 and 25°C) using DLS and a functional assay, in the optimal buffer for sample homogeneity and solubility (see Optimization of homogeneity and solubility section). Indeed, one may quite often realize this way that simple storage of the protein sample without further processing (for instance at 4°C) provides long enough stability for all down-stream experiments.

Many people also evaluate the thermal stability of their proteins in different buffers, using methods such as differential scanning fluorimetry (DSF, also known as thermal-shift assay) [57]: however, there is no clear correlation between thermodynamic and time stability of a protein, and it is therefore not straightforward to obtain insight about the long-term stability of a sample from its thermal stability analysis. On the contrary, thermodynamic stability generally correlates with rigidity [62], which is of particular importance when the downstream application is structural characterization (for instance by X-ray crystallography).

If a protein needs to be stored for an undetermined period, one can explore different methods (freezing with or without cryoprotectants, lyophilization,... [59-61]) and determine their effect on the properties of the sample using DLS and a functional assay. Of note, the best storage conditions may be largely different from the experimental conditions for downstream applications, so a preliminary desalting or dialysis might be needed before quality control.

Determination of protein sample reproducibility and lot-to-lot consistency

A fundamental principle of good laboratory practices is that experiments need to be reproduced and should thus

be reproducible, both within a laboratory and between research groups. During the lifetime of a project, it is therefore very likely that one will need to prepare more than a single sample of a given protein. Other groups might also need to prepare it independently in the frame of collaborations or comparability studies. Determining the robustness of one's production/purification process and its capacity to reproducibly deliver samples of equivalent quality is therefore all-important. However, once the quality of a purified protein sample has been fully assessed and optimized a first time, verification of lot-to-lot consistency does not necessarily require the repetition of the whole quality control work-flow (Figure 1B).

A very practical way to rapidly estimate the equivalence of protein lots is to verify the conformity of their "spectral signatures". The most straightforward is to compare UV-visible spectra which, as has been stressed above, contain a wealth of information beyond simple 280 nm absorbance. This may be profitably complemented by circular dichroism (CD) in the far-UV, which provides information about the global content of secondary structure elements in a protein [63,64]. Of note, contrary to a widespread belief, the presence of secondary structure elements in a protein ("foldedness") is not by itself a quality control criterium, especially as many proteins are either intrinsically disordered or contain unfolded segments in their native state. But differences between the CD spectra acquired for two different lots of the same protein (in the same buffer) may readily reveal divergences in folding that could correlate with differences in active concentration, especially if spectral similarity is analysed quantitatively rather than visually [65,66].

"Thermal denaturation signatures", determined by techniques such as CD or differential scanning calorimetry (DSC, [67]), can also be a very convenient and accurate way to determine the equivalence of protein lots, provided special attention is given to the equivalence of protein sample conditioning buffers. Indeed, differences between protein lots can translate into detectable differences in the global shape of their denaturation profiles [68].

Apart from spectral and thermal denaturation signatures, MS (for integrity), DLS (for homogeneity), analytical SEC (for both purity and homogeneity) and a functional assay are the most convenient and discriminating methods to assess the reproducibility and equivalence in quality of distinct protein lots.

Conclusion

In this review, we have attempted to cover all the aspects of protein quality control, from the necessary initial sample assessment to sample optimization. For each step, a set of relevant techniques has been suggested (Figure 1A). The first-line methods are essential and should be used systematically for a full quality control

assessment. Different complementary methods can be added depending on the protein sample peculiarities and quality control requirements. The suggested approaches for first line assessment include the “basic requirements for evaluating protein quality” that have been recently proposed [10], but go significantly beyond them. We also suggest a sequential experimental work-flow, to be followed as a check-list in order to optimize the time and effort spent on each sample (Figure 1B). This work-flow elaborates the protein quality control and storage optimization steps of the general protein production/purification pipeline [10]. Overall, this global synthetic step-by-step overview should hopefully lead to better protein samples and therefore to better chances of success in downstream applications. In line with community-based efforts that have been deployed in other fields like structural biology [69,70], proteomics and interactomics [71-74] or quantitative real-time PCR [75,76], research relying on purified proteins would gain significant reliability and credibility from the implementation of good practices, such as the systematic and transparent reporting of the results of purified protein quality control assessments, at least in the supplementary information sections of scientific publications.

Abbreviations

SDS-PAGE: Sodium dodecyl sulfate polyAcrylamide gel electrophoresis; MS: Mass spectrometry; MALDI: Matrix-assisted laser desorption/ionization; IEF: Iso-electric focusing; CE: Capillary electrophoresis; DLS: Dynamic light Scattering; SEC: Size Exclusion Chromatography; AFFF: Asymmetric Flow-Field flow fractionation; RI: Refractive index; SLS: Static light scattering; SPR: Surface plasmon resonance; CFCA: Calibration-Free Concentration Analysis; FTIR: Fourier Transform Infrared Spectroscopy; AAA: Amino acid analysis; CD: Circular dichroism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PL, BB and SH collected data and references for different parts of the review; BR and PE coordinated and assembled the manuscript. All authors read and approved the final manuscript.

Received: 25 September 2014 Accepted: 10 December 2014

Published online: 30 December 2014

References

- Leader B, Baca QJ, Golan DE: **Protein therapeutics: a summary and pharmacological classification.** *Nat Rev Drug Discov* 2008, **7**:21–39.
- Beck A, Wurch T, Bailly C, Corvaia N: **Strategies and challenges for the next generation of therapeutic antibodies.** *Nat Rev Immunol* 2010, **10**:345–352.
- Carter PJ: **Introduction to current and future protein therapeutics: a protein engineering perspective.** *Exp Cell Res* 2011, **317**:1261–1269.
- Rosenberg AS: **Effects of protein aggregates: an immunologic perspective.** *AAPS J* 2006, **8**:E501–507.
- Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, Dhe-Paganon S, Park H, Savchenko A, Yee A, Edwards A, Vincetelli R, Cambillau C, Kim R, Kim S-H, Rao Z, Shi Y, Terwilliger TC, Kim C-Y, Hung L-W, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, et al: **Protein production and purification.** *Nat Methods* 2008, **5**:135–146.
- Medrano G, Dolan MC, Condori J, Radin DN, Cramer CL: **Quality Assessment of Recombinant Proteins Produced in Plants.** In *Recombinant Gene Expression: Methods and Applications*. Edited by Lorence A. Totowa, NJ: Humana Press; 2012:535–564 [Methods in Molecular Biology, vol 824].
- Daviter T, Fronzes R: **Protein Sample Characterization.** In *Protein-Ligand Interactions: Methods and Applications*. Edited by Williams MA, Daviter T. Totowa, NJ: Humana Press; 2013:35–62 [Methods in Molecular Biology, vol 1008].
- De Marco A: **Minimal information: an urgent need to assess the functional reliability of recombinant proteins used in biological experiments.** *Microb Cell Fact* 2008, **7**:20.
- Buckle AM, Bate MA, Androulakis S, Cinquanta M, Basquin J, Bonneau F, Chatterjee DK, Cittaro D, Gräslund S, Gruszka A, Page R, Suppmann S, Wheeler JX, Agostini D, Taussig M, Taylor CF, Bottomley SP, Villaverde A, de Marco A: **Recombinant protein quality evaluation: proposal for a minimal information standard.** *Stand Genomic Sci* 2011, **5**:195–197.
- Ledenbiker M, Danieli T, de Marco A: **The Trip Adviser guide to the protein science world: a proposal to improve the awareness concerning the quality of recombinant proteins.** *BMC Res Notes* 2014, **7**:585.
- Walker JM: **SDS Polyacrylamide Gel Electrophoresis of Proteins.** In *Protein Protocols Handbook*. 3rd edition. Edited by Walker JM. Totowa, NJ: Humana Press; 2009:177–185.
- Fernandez-Patron C: **Zinc-Reverse Staining Technique.** In *Protein Protocols Handbook*. 3rd edition. Edited by Walker JM. Totowa, NJ: Humana Press; 2009:505–513.
- Chevallet M, Luche S, Rabilloud T: **Silver staining of proteins in polyacrylamide gels.** *Nat Protoc* 2006, **1**:1852–1858.
- Fernandez-Patron C, Hardy E, Sosa A, Seoane J, Castellanos L: **Double staining of coomassie blue-stained polyacrylamide gels by imidazole-sodium dodecyl sulfate-zinc reverse staining: sensitive detection of coomassie blue-undetected proteins.** *Anal Biochem* 1995, **224**:263–269.
- Hardy E, Castellanos-Serra LR: **“Reverse-staining” of biomolecules in electrophoresis gels: analytical and micropreparative applications.** *Anal Biochem* 2004, **328**:1–13.
- Irie S, Sezaki M, Kato Y: **A faithful double stain of proteins in the polyacrylamide gels with Coomassie blue and silver.** *Anal Biochem* 1982, **126**:350–354.
- Yan JX, Wait R, Harry RA, Westbrook JA, Wheeler CH, Dunn MJ: **A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption / ionization and electrospray ionization- mass spectrometry.** *Electrophoresis* 2000, **21**:3666–3672.
- Alba FJ, Bartolomé S, Bermúdez A, Daban J: **Fluorescent Labeling of Proteins and Its Application to SDS-PAGE and Western Blotting.** In *Protein Blotting and Detection: Methods and Protocols*. Edited by Kurien BT, Scofield RH. Totowa, NJ: Humana Press; 2009:407–416 [Methods in Molecular Biology, vol 536].
- Buxbaum E: **Fluorescent Staining of Gels.** In *Protein Electrophoresis: Methods and Protocols*. Edited by Kurien BT, Scofield RH. Totowa, NJ: Humana Press; 2012:543–550 [Methods in Molecular Biology, vol 869].
- Miller I, Crawford J, Gianazza E: **Protein stains for proteomic applications: which, when, why?** *Proteomics* 2006, **6**:5385–5408.
- Magdeldin S, Enany S, Yoshida Y, Xu B, Zhang Y, Zureena Z, Lokamani I, Yaoita E, Yamamoto T: **Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis.** *Clin Proteomics* 2014, **11**:16.
- Zhao SS, Chen DDY: **Applications of capillary electrophoresis in characterizing recombinant protein therapeutics.** *Electrophoresis* 2014, **35**:96–108.
- Haselberg R, de Jong GJ, Somsen GW: **CE-MS for the analysis of intact proteins 2010–2012.** *Electrophoresis* 2013, **34**:99–112.
- Glaser J: **Validity of nucleic acid purities monitored by 260 nm/280nm absorbance ratios.** *Biotechniques* 1995, **18**:62–63.
- Pace CN, Vajdos F, Fee L, Grimsley G, Gray T: **How to measure and predict the molar absorption coefficient of a protein.** *Protein Sci* 1995, **4**:2411–2423.
- Noble JE: **Quantification of Protein Concentration Using UV Absorbance and Coomassie Dyes.** In *Laboratory Methods in Enzymology: Protein Part A*. Edited by Lorsch J. Waltham, MS: Academic Press; 2014:17–26 [Methods in Enzymology, vol 536].
- Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Kelleher NL: **Analysis of intact protein isoforms by mass spectrometry.** *J Biol Chem* 2011, **286**:25451–25458.
- Witze ES, Old WM, Resing KA, Ahn NG: **Mapping protein post-translational modifications with mass spectrometry.** *Nat Methods* 2007, **4**:798–806.
- Debois D, Smargiasso N, Demeure K, Asakawa D, Zimmerman TA, Quinton L, De Pauw E: **MALDI In-Source Decay, from sequencing to imaging.** *Top Curr Chem* 2013, **331**:117–141.

30. Liu X, Dekker LJM, Wu S, Vanduijn MM, Luider TM, Tolić N, Kou Q, Dvorkin M, Alexandrova S, Vyatkin K, Paša-Tolić L, Pevzner P: **De novo protein sequencing by combining top-down and bottom-up tandem mass spectra.** *J Proteome Res* 2014, **13**:3241–3248.
31. Speicher KD, Gorman N, Speicher DW: **N-Terminal Sequence Analysis of Proteins and Peptides.** *Curr Protoc Protein Sci* 2009, **57**:11.10.1–11.10.31.
32. Zhang G, Annan RS, Carr S, Neubert T: **Overview of peptide and protein analysis by mass spectrometry.** *Curr Protoc Protein Sci* 2010, **62**:16.1.1–16.1.30.
33. Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Shen F, Yi J, Harding SE: **Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies.** *Biotechnol Genet Eng Rev* 2007, **24**:117–128.
34. Philo JS: **Is any measurement method optimal for all aggregate sizes and types?** *AAPS J* 2006, **8**:E564–571.
35. Schuck P: **Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling.** *Biophys J* 2000, **78**:1606–1619.
36. Leach SJ, Scheraga HA: **Effect of Light Scattering on Ultraviolet Difference Spectra.** *J Am Chem Soc* 1960, **82**:4790–4792.
37. Nominé Y, Ristriani T, Laurent C, Lefèvre J-F, Weiss E, Travé G: **A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein.** *Protein Eng* 2001, **14**:297–305.
38. Fekete S, Beck A, Veuthey J-L, Guillaume D: **Theory and practice of size exclusion chromatography for the analysis of protein aggregates.** *J Pharm Biomed Anal* 2014, **101**:161–173.
39. Diederich P, Hansen SK, Oelmeier S, Stolzenberger B, Hubbuch J: **A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography.** *J Chromatogr A* 2011, **1218**:9010–9018.
40. Barth HG, Saunders GD, Majors RE: **The State of the Art and Future Trends of Size-Exclusion Chromatography Packings and Columns.** *LC GC North Am* 2012, **30**:544–563.
41. Sala E, de Marco A: **Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography.** *Protein Expr Purif* 2010, **74**:231–235.
42. Gabrielson JP, Brader ML, Pekar AH, Mathis KB, Winter G, Carpenter JF, Randolph TW: **Quantitation of Aggregate Levels in a Recombinant Humanized Monoclonal Antibody Formulation by Size-Exclusion Chromatography, Asymmetrical Flow Field Flow Fractionation, and Sedimentation Velocity.** *J Pharm Sci* 2007, **96**:268–279.
43. Liu J, Andya JD, Shire SJ: **A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation.** *AAPS J* 2006, **8**:E580–589.
44. Sahin E, Roberts CJ: **Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms.** In *Therapeutic Proteins: Methods and Protocols*. Edited by Voynov V, Caravella JA. Totowa, NJ: Humana Press; 2012:403–423 [Methods in Molecular Biology, vol 899].
45. Ye H: **Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography-multiangle laser light scattering.** *Anal Biochem* 2006, **356**:76–85.
46. Zeder-Lutz G, Benito A, Van Regenmortel MH: **Active concentration measurements of recombinant biomolecules using biosensor technology.** *J Mol Recognit* 1999, **12**:300–309.
47. Sigmondsson K, Måsson G, Rice R, Beauchemin N, Obrink B: **Determination of active concentrations and association and dissociation rate constants of interacting biomolecules: an analytical solution to the theory for kinetic and mass transport limitations in biosensor technology and its experimental verification.** *Biochemistry* 2002, **41**:8263–8276.
48. Pol E: **The importance of correct protein concentration for kinetics and affinity determination in structure-function analysis.** *J Vis Exp* 2010, **37**:2–8.
49. England P, Brégégère F, Bedouelle H: **Energetic and kinetic contributions of contact residues of antibody D1.3 in the interaction with lysozyme.** *Biochemistry* 1997, **36**:164–172.
50. Grimsley G, Pace CN: **Spectrophotometric determination of protein concentration.** *Curr Protoc Protein Sci* 2003, **33**:3.1.1–3.1.9.
51. Etzion Y, Linker R, Cogan U, Shmulevich I: **Determination of protein concentration in raw milk by mid-infrared fourier transform infrared/attenuated total reflectance spectroscopy.** *J Dairy Sci* 2004, **87**:2779–2788.
52. Rutherford SM, Gilani GS: **Amino Acid Analysis.** *Curr Protoc Protein Sci* 2009, **58**:11.9.1–11.9.37.
53. Saraswat M, Musante L, Ravid A, Shortt B, Byrne B, Holthofer H: **Preparative purification of recombinant proteins: current status and future trends.** *Biomed Res Int* 2013, **2013**: Article ID 312709, 18 pages. doi:10.1155/2013/312709
54. Lebendiker M, Danieli T: **Production of prone-to-aggregate proteins.** *FEBS Lett* 2014, **588**:236–246.
55. Wang J, Matayoshi E: **Solubility at the molecular level: development of a critical aggregation concentration (CAC) assay for estimating compound monomer solubility.** *Pharm Res* 2012, **29**:1745–1754.
56. Jancarik J, Pufan R, Hong C, Kim SH, Kim R: **Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins.** *Acta Crystallogr D Biol Crystallogr* 2004, **60**:1670–1673.
57. Boivin S, Kozak S, Meijers R: **Optimization of protein purification and characterization using Thermofluor screens.** *Protein Expr Purif* 2013, **91**:192–206.
58. Cao E, Chen Y, Cui Z, Foster PR: **Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions.** *Biotechnol Bioeng* 2003, **82**:684–690.
59. Carpenter JF, Manning MC, Randolph TW: **Long-term storage of proteins.** *Curr Protoc Protein Sci* 2002, **27**:4.6.1–4.6.6.
60. Patro SY, Freund E, Chang BS: **Protein formulation and fill-finish operations.** *Biotechnol Annu Rev* 2002, **8**:55–84.
61. Simpson RJ: **Stabilization of proteins for storage.** *Cold Spring Harb Protoc* 2010, **5**: doi:10.1101/pdb.top79.
62. Jaenicke R: **Stability and stabilization of globular proteins in solution.** *J Biotechnol* 2000, **79**:193–203.
63. Greenfield NJ: **Using circular dichroism spectra to estimate protein secondary structure.** *Nat Protoc* 2006, **1**:2876–2890.
64. Li CH, Nguyen X, Narhi L, Chennamalil L, Towers E, Muzammil S, Gabrielson J, Jiang Y: **Applications of Circular Dichroism (CD) for structural analysis of proteins: qualification of near- and far-UV CD for protein higher order structural analysis.** *J Pharm Sci* 2011, **100**:4642–4654.
65. Ravi J, Rakowska PD, Garfagnini T, Baron B, Charlet P, Jones C, Milev S, DeSa LJ, Plusquellic D, Wien F, Wu L, Meuse CW, Knight AE: **International comparability in spectroscopic measurements of protein structure by circular dichroism: CCQM-P59.1.** *Metrologia* 2010, **47**:631–641.
66. Teska BM, Li C, Winn BC, Arthur KK, Jiang Y, Gabrielson JP: **Comparison of quantitative spectral similarity analysis methods for protein higher-order structure confirmation.** *Anal Biochem* 2013, **434**:153–165.
67. Johnson CM: **Differential scanning calorimetry as a tool for protein folding and stability.** *Arch Biochem Biophys* 2013, **531**:100–109.
68. Wen J, Arthur K, Chennamalil L, Muzammil S, Gabrielson J, Jiang Y: **Applications of differential scanning calorimetry for thermal stability analysis of proteins : qualification of DSC.** *J Pharm Sci* 2012, **101**:955–964.
69. Berman HM, Kleywegt GJ, Nakamura H, Markley JL: **How community has shaped the Protein Data Bank.** *Structure* 2013, **21**:1485–1491.
70. Read RJ, Adams PD, Arendall WB, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütke T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, Zwart PH: **A new generation of crystallographic validation tools for the Protein Data Bank.** *Structure* 2011, **19**:1395–1412.
71. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJ, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping P, Seymour SL, Souida P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR 3rd, Hermjakob H: **The minimum information about a proteomics experiment (MIAPE).** *Nature Biotechnol* 2007, **25**:887–893.
72. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin AC, Superti-Furga G, Greenblatt J, Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las RJ, Prieto C, Perreau VM, Hogue C, Mewes HW, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H: **The minimum information required for reporting a molecular interaction experiment (MIMIx).** *Nature Biotechnol* 2007, **25**:894–898.
73. Martínez-Bartolomé S, Binz PA, Albar JP: **The Minimal Information about a Proteomics Experiment (MIAPE) from the Proteomics Standards Initiative.** In *Plant Proteomics: Methods and Protocols*. Edited by Jorin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S. Totowa, NJ: Humana Press; 2014:765–780 [Methods in Molecular Biology, vol 1072].

74. Eisenacher M, Schnabel A, Stephan C: **Quality meets quantity - quality control, data standards and repositories.** *Proteomics* 2011, **11**:1031–1036.
75. Johnson G, Nour AA, Nolan T, Huggett J, Bustin S: **Minimum information necessary for quantitative real-time PCR experiments.** In *Quantitative Real-Time PCR: Methods and Protocols*. Edited by Biassoni R, Raso A. Totowa, NJ: Humana Press; 2014:5–17 [*Methods in Molecular Biology*, vol 1160].
76. Bustin SA, Benes V, Garson J, Hellems J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley G, Wittwer CT, Schjerling P, Day PJ, Abreu M, Aguado B, Beaulieu JF, Beckers A, Bogaert S, Browne JA, Carrasco-Ramiro F, Ceelen L, Ciborowski K, Cornillie P, Coulon S, Cuypers A, De Brouwer S, De Ceuninck L, De Craene J, De Naeyer H, De Spiegelaere W, et al: **The need for transparency and good practices in the qPCR literature.** *Nat Methods* 2013, **10**:1063–1067.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

