CrossMark

Guidelines to reach high-quality purified recombinant proteins

Carla Oliveira[1] · Lucília Domingues[1]

Abstract The final goal in recombinant protein production is to obtain high-quality pure protein samples. Indeed, the successful downstream application of a recombinant protein depends on its quality. Besides production, which is conditioned by the host, the quality of a recombinant protein product relies mainly on the purification procedure. Thus, the purification strategy must be carefully designed from the molecular level. On the other hand, the quality control of a protein sample must be performed to ensure its purity, homogeneity and structural conformity, in order to validate the recombinant production and purification process. Therefore, this review aims at providing succinct information on the rational purification design of recombinant proteins produced in *Escherichia coli*, specifically the tagging purification, as well as on accessible tools for evaluating and optimizing protein quality. The classical techniques for structural protein characterization—denaturing protein gel electrophoresis (SDS-PAGE), size exclusion chromatography (SEC), dynamic light scattering (DLS) and circular dichroism (CD)—are revisited with focus on the protein and their main advantages and disadvantages. Furthermore, methods for determining protein concentration and protein storage are also presented. The guidelines compiled herein will aid preparing pure, soluble and homogeneous functional recombinant proteins from the very beginning of the molecular cloning design.

Introduction

Recombinant protein production (RPP) has been increasingly used in laboratorial research for obtaining recombinant proteins for biophysical and structural studies (Vedadi et al. 2010), diagnostic and therapeutic purposes (Jozala et al. 2016), as well as emerging applications, such as smart materials (Hollingshead et al. 2017). The RPP field constitutes a multi-billion dollar market since a significant part of the main biotechnological market products are recombinant proteins. Namely, the total market sales from microbial recombinant products have reached approximately $50 billion in 2016, representing one third of the total sales of biopharmaceuticals (Jozala et al. 2016). Regardless the final protein application, high-quality protein samples must be obtained upon the RPP process, which could fulfil established purity and conformational requirements. Purification and characterization of recombinant proteins can be demanding, expensive and time-consuming, but can determine protein quality. That is, the successful application of a recombinant protein depends, to a great extent, on its efficient downstream processing. This comprises protein purification, quality validation, quantification and storage. The design of a rational protein purification strategy should be the first step in the overall RPP strategy. A poor purification design may result in misfolded or heterogeneous protein samples due to the interference of sequence additions, such as tags or extra amino acids resulting from the cloning procedure, and/or absence of refining steps in the purification procedure. Thus, adequate structural studies must be conducted to access the quality of the purified

✉ Lucília Domingues
luciliad@deb.uminho.pt

[1] CEB - Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

recombinant proteins and/or to aid optimizing their purity, homogeneity and solubility.

The RPP field has attracted a raising interest as indicated by the number of new molecular tools and methods described in recent years (e.g. Costa et al. 2014; Oliveira et al. 2015). Protein purification and structural characterization have already been the focus of some reviews (e.g. Manta et al. 2011; Saraswat et al. 2013; Yadav et al. 2016). Moreover, consensus RPP protocols (Gräslund et al. 2008) and workflows for quality control have been proposed (e.g. Lebendiker et al. 2014; Raynal et al. 2014) to instruct researchers in the production of soluble and reliable recombinant products. However, the subjects of protein purification and quality assessment are found separately, while both are strictly linked. Furthermore, both subjects are essential for those who are working in the RPP field and thus would be preferably found in a single document. Thus, this review intends to get relevant information altogether, by providing a succinct updated summary on the purification and quality evaluation of recombinant proteins. For more detailed information on the different referenced techniques, the reader is directed to specific reviews on the topic. Specifically, we will give key guidelines for designing a rational purification strategy and choosing accessible methods for characterizing recombinant proteins in order to obtain high-quality samples. The core of the review will be divided into three main sections, namely (i) protein purification strategies, (ii) protein quality assessment: structural characterization and (iii) optimization of protein stability, quantification and storage. Important aspects in the molecular design of the purification procedure will be highlighted, such as how to introduce removable protein tags in gene cloning and the use of bioinformatics to predict protein properties and structure, to help choosing fusion partners and where to attach them in the protein. In sum, a concise review covering essential aspects of purification and refining of recombinant protein samples will be provided for researchers starting in the area as well as experienced researchers searching for compiled and updated information.

Protein purification strategies

The purification of recombinant proteins must be carefully designed not only because this is the most expensive step in RPP but also due to its significant impact on protein function/application. There is no single or simple way to purify all kinds of proteins because of their diversity and different properties (Costa et al. 2014). Furthermore, the purification procedure depends on the expression system, and namely, if the protein is produced inside or outside de host cells. In yeast, such as *Pichia pastoris*, production is typically directed into the extracellular medium, to allow post-translational

modifications (e.g. glycosylation), where low levels of native proteins are found. Thus, in this case, purification is quite simple; a single purification step may be sufficient (Oliveira et al. 2008). In *Escherichia coli*, purification is more complex because recombinant proteins are mainly produced inside the cells, even though the expressed protein can comprise up to 50% of the total cellular protein (Francis and Page 2010). The purification strategy also depends on the purity level required. Efficient purification of a target protein from a crude cell extract is not always straightforward, and sometimes, more than one purification step is necessary. As such, fusion tag technology has been employed to facilitate purification, by reducing purification steps to a minimum, while increasing yield and purity. The main difficulty is to decide which and how many tags should be employed, where to attach them in the protein and whether or not they should be removed from the protein. The main guidelines to design a rational tagging purification strategy are discussed below.

Overall fusion strategy: molecular design of the purification procedure

The purification strategy should be designed together with the gene expression procedure (Saccardo et al. 2016). There is no universal fusion tag to purify any recombinant protein, and many times, it requires experimental identification (Lebendiker and Danieli 2014). However, the design of a rational purification strategy can increase the chances of success. The choice of the fusion partner often relies on costs and tools available at each lab. However, tag type, size and location can be the most important criteria of selection.

(i) *Tag type.* The most used purification tags for protein purification are the affinity tags (see "Tag-dependent purification" section). Affinity tags can deliver different purposes. The His-tag is the most popular affinity tag either as single tag or as a handle in non-traditional tandem affinity purification (TAP) tags (i.e. two affinity tags fused in tandem) (Yadav et al. 2016). Its utility in protein detection and crystallization has also been well established (Costa et al. 2014). However, contrary to a general belief, it can interfere with protein structural and/or functional properties (e.g. Noirclerc-Savoye et al. 2015; Chen et al. 2015). For solving problems of insolubility, affinity tags that also work as solubility enhancers are used (see "Tag-dependent purification" section) (Costa et al. 2014). New computational tools for predicting protein solubility from the primary sequence have been recently described, such as ccSOL (Agostini et al. 2014), Protein-Sol (Hebditch et al. 2017) and PON-Sol (Yang et al. 2016) (Table 1). Of note, only ~ 20% of all recombinant proteins expressed in bacteria are

produced in soluble form, and this percentage is reduced to ~10% for mammalian-sourced proteins (Gräslund et al. 2008). TAP tags are employed when high purity levels of protein are required, as is the case of therapeutic and biomedical applications (Zou et al. 2017). Affinity purification tags suitable for in vivo studies are also available (HaloTag Technology, England et al. 2015). The advantages and limitations of commonly used affinity tags have been previously discussed with great detail (e.g. Costa et al. 2014; Yadav et al. 2016).

(ii) *Tag size.* Small tags are typically chosen for being less prone to impact on folding, biological activity and immunogenicity of the target protein. However, larger tags can provide higher protein production levels due to their strong translational initiation signals. Moreover, if the target protein has a small size, e.g. peptides, large tags are chosen to prevent host expression problems (Guerreiro et al. 2008; Ramos et al. 2010, 2013). Tag size also depends on the overall conformation of the protein (see (iii) *Tag location*). Independently of their size, tags can be easily introduced in gene cloning using PCR methods (Aguiar et al. 2017; Oliveira et al. 2017). Basically, the base pair sequence of tags are included in conventional primers (small tags) or megaprimers (large tags), following general rules of primer design (Oliveira et al. 2017), which are subsequently assembled and amplified, together with the target gene, in single or multi-step PCR protocols, respectively. Tag removal is often necessary because tags, the large ones in particular, may interfere with the structure and function of the protein. Tags are generally removed from the target protein using specific proteases, whose enzymatic sequence site is placed between the tag and the protein. Thus, for further tag cleavage, the base pair sequence coding this site (most protease recognition sites are very small, ranging from four to eight amino acids) can be placed in primers (5′-3′) downstream the tag sequence (Costa et al. 2013a). Specific codons for chemical cleavage sites can be placed instead, such as proline and aspartate triplets for formic acid cleavage (Ramos et al. 2010, 2013). The proteolysis susceptibility (Table 1) of the target protein to the chemical should be previously estimated.

(iii) *Tag location.* A favourable exposition of the N-terminal and/or C-terminal of the protein for tagging should be evaluated ideally by analysis of its tertiary structure (if already deposited at the RSCB Protein Data Bank (PDB)) or by homology-based modelling using the resolved structure of closely related proteins as structural template (SWISS MODEL, Biasini et al. 2014) (Table 1). In case both ends are hidden (i.e. faced inwards the structure), it does not necessarily mean that the protein cannot be tagged. Fusion protein linkers can be used to improve tag accessibility for purification

(Chen et al. 2013). Larger tags can also be used instead of smaller ones for the same reason. For instance, large tags facilitated His-tag purification of tetrameric plant lectin frutalin from *E. coli*, while significantly improving its solubility and stability (Oliveira et al. 2014a). If a free (native) N-terminal is determinant for protein activity, the C-terminal should be chosen for tag location. Alternatively, as C-terminal tagging is more favourable for efficient expression, tag can be placed at this end and cleaved with a protease that leaves none (e.g. factor Xa) or few additional residues (e.g. TEV) at the N-terminal of the protein (most proteases cleave at or near the C-terminus of their own recognition sites). Chemical cleavage (Ramos et al. 2010, 2013) and intein self-cleavage (Shi et al. 2017) also leave few or none extra amino acids and thus can be considered for a more cost-effective tag removal procedure. Tag removal must be also carefully designed as it has many practical difficulties associated (for a review, see, e.g. Yadav et al. 2016). Anyway, the effect of tag and its cleavage on protein properties should be evaluated by functional and structural assays (see "Protein quality assessment: structural characterization" section).

Tag-dependent purification

Affinity tagging has boosted the efficient purification of recombinant proteins. The sizes of the affinity tags range from < 1 kDa (e.g. 3–10× His-tag) to ~43 kDa (MBP (maltose binding protein)) (Yadav et al. 2016). They can be attached either to the N-terminal or C-terminal of the protein, or fused in tandem (known as TAP tags), usually in a dual format. TAP tag technology allows efficient protein purification from cell extracts (i.e. very contaminated samples) in two consecutive steps. Compared with single-step purification, TAP significantly reduces non-specific background and isolates protein complexes with higher purity (Li 2011). Although initially developed in yeast for protein-protein interaction studies, TAP tags have been efficiently modified in various expression systems for protein purification, including yeast, bacteria, insect and mammalian cells and plants (Yadav et al. 2016). Emerging multi-TAP combinations are being developed to amplify the versatility and downstream applications (Yadav et al. 2016). For example, very recently, the TAP tag His-SBP (streptavidin-binding peptide, ~5 kDa) was used to purify efficiently from *E. coli* the heterodimeric isoform CK-MB of the human creatine kinase, an important biomarker of myocardial injury (Zou et al. 2017). Interestingly, after simultaneous expression of CK-*M* and CK-*B* genes from a dual expression vector, subsequent two-step purification and tag removal with factor Xa, the recombinant product exhibited equal parts of M and B isoforms in SDS-PAGE and presented

Table 1 List of free computational tools frequently used for protein analysis

Application	Tool name	Web link
Protein properties		
Amino acid composition and mass	ProtParam	http://web.expasy.org/protparam
Isoelectric point	Compute pI/Mw	http://web.expasy.org/compute_pi
Hydrophobicity	Hydrophobicity Plotter (Innovagen)	http://pepcalc.com/
Cysteine residues		
Disulphide bond connectivity	DiANNA	http://clavius.bc.edu/~clotelab/DiANNA/
Proteolysis susceptibility		
Cutting sites for enzymes and chemicals	PeptideCutter	http://web.expasy.org/peptide_cutter/
Solubility		
Expression in *E. coli*	Protein-Sol	http://protein-sol.manchester.ac.uk
	ccSOL	http://s.tartaglialab.com/page/ccsol_group
	PON-Sol	http://structure.bmc.lu.se/PON-Sol
	Recombinant Protein Solubility Prediction	http://www.biotech.ou.edu/
Aggregation propensity		
Prediction of "hot spots" of aggregation in polypeptides	AGGRESCAN	http://bioinf.uab.es/aggrescan/
ThermoFluor		
Tm curves analysis	Meltdown	https://github.com/C3-CSIRO/Meltdown
Circular dichroism		
CD data analysis	DICHROWEB	http://dichroweb.cryst.bbk.ac.uk/html/home.shtml
	K2D2	http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2//
	K2D3	http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/
3D structure		
Structure modelling	I-TASSER	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
	YASARA	www.yasara.org
	SWISS MODEL	https://swissmodel.expasy.org/
Structural parameters	PDBparam	http://www.iitm.ac.in/bioinfo/pdbparam/

high catalytic activity and good commutability with an identical standard commercial enzyme (Zou et al. 2017). The subject of fusion tags has already been well described in many excellent reviews (see, e.g. Costa et al. 2014; Yadav et al. 2016). Nevertheless, novel tags, from natural origin (Li et al. 2016) or rationally designed (Morris et al. 2016), are continually being reported in the literature. Among the affinity tags recently described, interesting alternatives to the His-tag have been presented (Vargas-Cortez et al. 2016; Cantu-Bustos et al. 2016). Namely, the SmbP and CusF tags, small metal-binding proteins of 9.9 kDa isolated from *Nitrosomonas europaea* and *E. coli*, respectively, could be used to purify fluorescent model proteins from *E. coli* at yields higher than that obtained using the His-tag. Furthermore, tags increased recombinant protein solubility compared to the levels obtained using the common fusion proteins MBP and GST (Vargas-Cortez et al. 2016; Cantu-Bustos et al. 2016). Some fusion tags for protein

solubility that also work as purification tags have been likewise reported in recent years. For example, the novel Fh8 tag, a small (8 kDa) calcium-binding protein isolated from *Fasciola hepatica*, has shown outstanding combined solubility and purification abilities in *E. coli* (Costa et al. 2014). Namely, cleaved proteins from Fh8 fusions were soluble and obtained in similar or higher amounts than proteins cleaved from other partners as Trx, NusA or MBP (Costa et al. 2013a). Moreover, Fh8 allowed purification of fused proteins by calcium-dependent hydrophobic interaction chromatography (HIC) at efficiencies identical to those of the His-tag (Costa et al. 2013b). Of note, the Fh8-HIC methodology presented also the advantage of being compatible with His-tag purification, thus, allowing a dual protein purification strategy that can be used sequentially, complementing each other, to obtain a more pure protein when desired (Costa et al. 2014). More recently, a novel self-cleavable tag called Zbasic-intein

improved the solubility of recombinant cytokine human interleukin-15 (IL-15) in *E. coli* and facilitated its purification by cation exchange chromatography (Shi et al. 2017). Many aggregation tags, directed for column-free purification, have also emerged in recent years, with yield and purity generally comparable to that of the His-tag (for a review, see Lin et al. 2015). These tags induce the formation of aggregates (during or after expression) when fused to a target protein or peptide, and upon separation from soluble impurities, the target protein or peptide is subsequently released via a cleavage site (Lin et al. 2015). Although very promising with respect to large-scale cost savings in resin and columns, they have not been extensively used because they present multiple shortcomings (Lin et al. 2015), such as co-aggregation and entrainment of impurities during precipitation and centrifugation, as reported in Shi et al. (2013).

Tag-free purification: optimization of purity and homogeneity

Although affinity purification has several recognized advantages (Costa et al. 2014), it has as major disadvantage the fact of being incapable of solving problems of protein heterogeneity, as charge and size variability. Thus, conventional chromatographic purification methods, based on protein properties such as size (size exclusion chromatography (SEC)), charge (ion exchange chromatography (IEX)) and hydrophobicity (HIC) (Saraswat et al. 2013), are often coupled downstream the affinity purification procedure for sample refining (Fig. 1). Among them, SEC (also known as gel filtration) is a well-established technique to remove protein aggregates, and thus to purify protein to size homogeneity, while providing information regarding the correct oligomeric structure of the protein. SEC is currently the standard separation technique for the separation/quantification of protein dimers, trimers and other oligomers, whose structures are essential for the activity of many proteins. It is also routinely used for desalting and buffer exchange of protein samples. Its main advantage is the mild (native) conditions that allow for the characterization of the protein with minimal impact on the conformational structure, and thus preserving its biological activity. The main disadvantage is the dilution of the protein sample during separation, which may alter equilibria between oligomeric species, while a concentration step may be required for downstream applications, which may induce protein precipitation (Raynal et al. 2014). Theory and practice of SEC for the analysis of protein aggregates have been recently reviewed, including advances in the improvement of throughput and resolution (Brusotti et al. 2017; Fekete et al. 2014). To ensure protein homogeneity, SEC should be the last step of the tagging purification strategy. Nonetheless, emerging methods, such as mixed-mode chromatography (based mainly in electrostatic and hydrophobic interactions) (Zhang and Liu 2016),

has attracted rising attention in recent years for tag-free purification of correctly folded recombinant proteins with high purity level, without the need for further purification or polishing steps (Gieseler et al. 2017).

Protein quality assessment: structural characterization

Structural characterization of proteins can be a very demanding and time-consuming task. This is because it requires large amounts of pure protein sample, appropriated equipment and skilled labour. The level of characterization depends on the final application of the protein, and most of the times, full characterization is not necessary. In fact, the best method to assess the correct conformation of a recombinant protein is to test its biological activity (Manta et al. 2011). However, even if the protein has the expected activity, characterization of a purified protein in some detail is a mean to assess and ensure its quality and to verify lot-to-lot consistency. There are three important aspects that must be carefully evaluated: purity, homogeneity and structural conformity. There are several methods that can address each one of these issues; the most accessible are discussed below.

Protein purity

Purity is the first quality that must be evaluated in a protein sample before any further experiment (Fig. 1). Common contaminants of a recombinant protein product can result from the culture medium, or can be incorporated from the extraction and purification procedures (e.g. host contaminants). Thus, a hypothetically pure protein sample can be in fact contaminated with many different molecules such as nucleic acids, endotoxins and/or other proteins. Nucleic acid contamination can be easily detected by UV-visible spectroscopy. A high 260/280-nm absorbance ratio indicates this contamination, whereas a ratio close to 0.57 corresponds to a pure protein sample (Glasel 1995). Endotoxin contamination is critical for therapeutic applications; conventional and emerging methods for its detection have been recently reviewed (Dullah and Ongkudon 2017). To evaluate contamination by other proteins, denaturing protein gel electrophoresis (SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis)) is normally used. Purity level of a protein sample by the criteria of gel electrophoresis is well accepted within the research community, being considered a basic requirement in quality control (Lebendiker et al. 2014). In fact, SDS-PAGE combined with the classical silver nitrate staining, or other staining method with similar sensitivity, is the most modest but reliable technique to certify protein purity for most protein applications. Silver staining

Fig. 1 Schematic representation of the purification and quality assessment of recombinant proteins, from molecular design to application. The purification of recombinant proteins is typically conducted via a fusion partner (tag-dependent purification), being the affinity tags the most widely used. The asterisk indicates that tag may be cleaved before protein refining (tag-free purification). The basic assays conducted after protein purification (purity and homogeneity evaluation) are within a box coloured in green. Structural conformity assays for studying protein folding are coloured in orange. Assays for optimizing protein stability under storage and functional assay conditions are coloured in pink. After this optimization step, protein homogeneity must be verified. Dashed lines represent steps that may require optimization
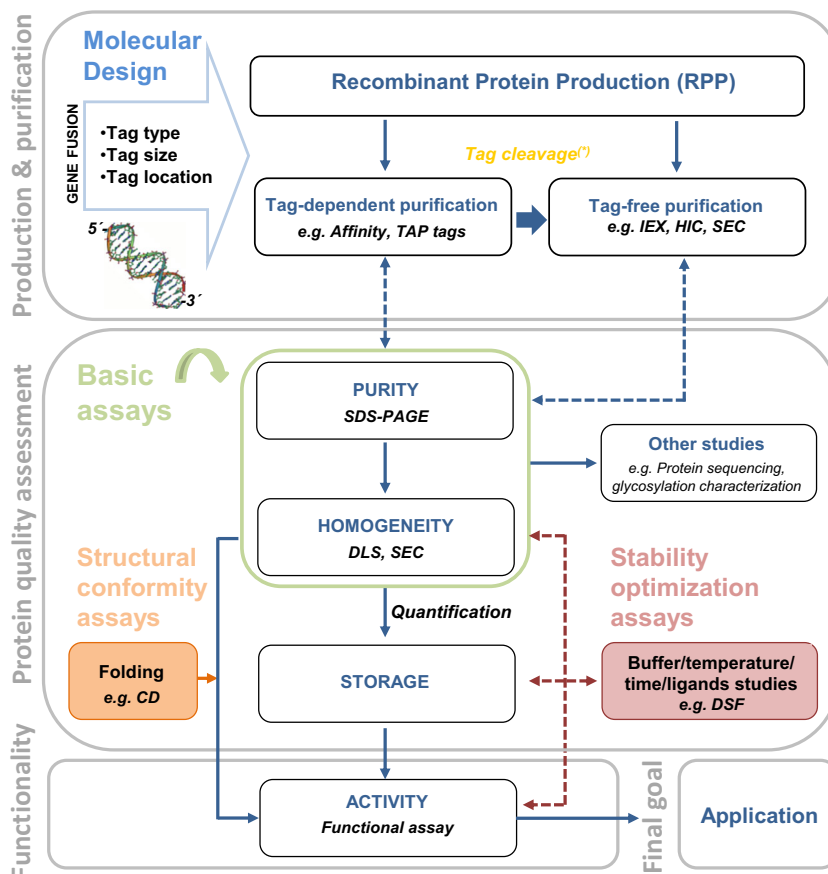
has an excellent sensitivity because it can detect protein amounts as low as 1 ng (Chevallet et al. 2006). Simplified and optimized protocols, compatible with downstream processing such as mass spectrometry (MS) and quantitative response, can be found in the literature (e.g. Rabilloud 2012). Hence, SDS-PAGE allows estimating the approximate concentration of a protein, as well as its molecular weight (MW), upon gel calibration with suitable protein standards. However, the presence of a single band with the expected MW does not necessarily mean that the protein has been purified to homogeneity. This must be evaluated under native conditions using, e.g. native PAGE or other appropriated methods (see "Protein homogeneity" section). Zymograms based on native PAGE or SDS-PAGE (from which SDS is removed, followed by protein renaturation) allows assessing in situ the referenced activity of the recombinant enzymes, with detection of active glycosylated isoforms, thus providing information on homogeneity and structural conformity (Magalhães et al. 2014; Ribeiro et al. 2010). Another drawback of protein gel electrophoresis is that it cannot separate contaminating proteins with a MW close to the MW of the target protein; this should be done using a technique with higher sensitivity, such as capillary electrophoresis (De Jong et al. 2016). On the other hand,

SDS-PAGE may detect protein degradation (smeared bands of lower MW than the target protein), glycosylation (using specific staining) and, in some cases, oligomerization. For example, recently, SDS-PAGE gel stained with Coomassie brilliant blue R-250 revealed monomeric (19 kDa), dimeric (38 kDa) and oligomeric (> 100 kDa) forms of a human growth factor (VEGF) recombinantly produced from *E. coli*, by analysis of a protein sample prepared under non-reducing conditions (Nguyen et al. 2016). SDS-PAGE can also be used to purify a protein (through band excision) from a partly purified protein sample for analysis of its primary structure by Edman degradation (N-terminal sequencing) (Oliveira et al. 2008) or MS-based methods (Zhang et al. 2014), for detecting desired or undesired small proteolysis events, or other modifications, undetectable by SDS-PAGE, in order to evaluate protein integrity.

Protein homogeneity

A protein sample is homogeneous if all molecules present have the same size and, expectedly, are fully folded in its native state. Soluble low or high MW aggregates are formed by non-covalent association of two or more polypeptide

chains, which may or may not retain their native fold (Murphy and Roberts 2013). Aggregation depends on intrinsic (i.e. structural features) or extrinsic factors (environment in which protein is present, processing conditions, etc.) (Wang et al. 2010). Aggregation propensity of a protein can be evaluated by computational approaches using its primary structure (Pallares and Ventura 2017) (Table 1). Such protein complexes may present reduced or no biological activity, besides other side effects (e.g. immunogenicity), and thus must be accurately detected. There are some straightforward methods for detecting protein aggregates, such as UV-visible and fluorescence spectroscopies or dynamic light scattering (DLS) (Khan and Kumar 2017). DLS, also known as photon correlation spectroscopy, has the advantage of combined sensitivity, reliability and broad applicability. Practical applications of DLS for evaluating the presence of soluble high-order assemblies and protein aggregates and studying protein interactions with other proteins, nucleic acid and small molecules (e.g. endotoxins) have been recently reviewed (Stetefeld et al. 2016; Minton 2016). DLS has also many other attractive advantages, such as it is a rapid technique, it requires low amounts of native samples, with simple or no previous preparation, it is suitable for studying macromolecules with a wide MW range, and it can detect trace amounts of high MW aggregates. Moreover, a large number of samples may be screened at once if using high-throughput instrumental analysis (e.g. He et al. 2010). The more suitable analysis method for studying protein aggregation is an *intensity distribution method*, as high MW aggregates will disproportionately scatter more light relative to smaller molecules enabling detection even if present at a relatively low concentration (Stetefeld et al. 2016). Distribution is plotted against an apparent hydrodynamic radius, i.e. radius of a hypothetical sphere that diffuses at the same rate as the particle under investigation (Stetefeld et al. 2016), which can be used to estimate the MW of the target molecules using the instrument software. However, autocorrelation functions must be carefully interpreted and, preferably, presented together with processed data (Minton 2016). DLS can also be employed for studying the stability of the proteins over time and/or at different temperatures, in diverse buffers, also in high-throughput format—differential static light scattering (DSLS) (Senisterra et al. 2012). However, DLS has also a number of limitations (Stetefeld et al. 2016): measurements are very sensitive to temperature and solvent viscosity (temperature must be kept constant, and solvent viscosity must be known for a reliable experiment); it is a low-resolution method that often cannot distinguish close quaternary structures (e.g. monomer from dimer); this distinction should be done by SEC (see "Tag-free purification: optimization of purity and homogeneity" section); it is not very reliable nor reproducible for measuring the MW (for that, alternate methods such as SLS-based methods can be used); it is restricted to transparent samples;

large contaminating particles affect measurements (sample-holding cuvette must be cleaned thoroughly, and sample must be filtered prior to measurements); the signal depends on the size and concentration of macromolecules, and thus, optimization of a range of concentrations may be required to obtain reliable measurements. Finally, because the technique is so simple (due to a simple set-up and fully automated measurement), there is a high risk of casual users to over-interpret DLS quantitative results (Minton 2016). Nevertheless, qualitative analysis of DLS results is sufficient for the application discussed herein.

Structural conformity

The study of secondary and tertiary structure of proteins is a valuable complementary approach in quality control to validate protein folding (Fig. 1). The principles and applications of five spectroscopic techniques suitable for monitoring protein conformational changes have been recently reviewed in Wang et al. (2017), namely Fourier transform infrared (FTIR), Raman, circular dichroism (CD), fluorescence and ultraviolet (UV) spectroscopies. CD is widely used in the RPP field for determining the secondary structures and folding properties of recombinant proteins, and in particular, the effects of mutations and ligands, and also fusion tags on protein and polypeptide stability (Healey et al. 2017; Zvonova et al. 2017). It is also routinely used for studying the unfolding and folding of proteins as a function of temperature. For example, this CD valence was employed to characterize the broad bactericidal activity of a phage endolysin produced in *E. coli* (Oliveira et al. 2014b). CD analysis demonstrated the exceptional ability of the protein to refold into its original conformation upon thermal denaturation up to 75 °C, thus explaining how it can retain activity after exposure of temperatures higher than its melting temperature of 44 °C (Tm, the midpoint of the unfolding transition). The great advantage of CD over the other techniques is that it is a fast technique that requires low amounts of protein (less than 20 μg, in the concentration range of 0.005–5 mg/ml, depending on the path length of the CD cuvette) (Greenfield 2006; Wang et al. 2017). However, the major difficulty of CD is the preparation of a proper protein sample. Samples should be highly clear with no insoluble protein aggregates present, as these will cause artefacts (Kelly et al. 2005). Moreover, protein concentration should be accurately determined for a reliable analysis. In fact, the trickiest part of obtaining high-quality CD data is the correct determination of protein concentration. Dye-based methods that have variable responses should not be used for this purpose (see "Protein quantification" section). Protein for CD analysis should be at least 95% pure, desalted (or dialysed) into the buffer and filtered (to reduce light scattering) immediately before analysis. Importantly, concentration must be checked after any protein manipulation. Buffers should be transparent

and free of any optically active material. It is mandatory to ensure the compatibility of buffers and additives (e.g. glycerol ≤ 20%) with the CD technique. Ideally, they should only make small contributions to the overall absorbance of the sample over the wavelength range of interest (Kelly et al. 2005). More details on protein and buffer preparation for CD can be found enclosed in specialized protocols. For example, the protocol by Greenfield (2006) is very useful because it also contains a troubleshooting guide. Besides the easy access to dedicated literature and intuitive computational tools (Table 1), acquirement and analysis of CD data require experienced knowledge.

Optimization of protein stability, quantification and storage

Stability studies

Stability studies consist in the evaluation of the protein susceptibility to the action of denaturant or degrading agents, such as temperature, oxidants or proteases. Thermal denaturation-based methods have increasingly become the methods of choice for screening proteins against several compounds and conditions that may stabilize proteins (Senisterra et al. 2012). This is particularly important for structural or functional assays, which are dependent on very specific conditions. Protein thermal denaturation can be analysed by different techniques, depending on the structural level of analysis, such as CD (see "Structural conformity" section). However, there has been a rising interest in techniques adaptable into high-throughput format for easy multiple parametric screening. One of the most used is the fluorescence-based thermal shift (ThermoFluor) assay (i.e. DSF (differential scanning fluorimetry)) (e.g. Boivin et al. 2013; Reinhard et al. 2013; Tileva et al. 2017). This method uses an environmentally sensitive dye, mainly Sypro Orange, to monitor the thermal stability of a pure protein under different buffer conditions. The principle of the technique is based on the detection of changes in the exposure of the hydrophobic core of the protein upon heat denaturation. Dye becomes fluorescent only when it intercalates into a hydrophobic pocket of the unfolded protein. Real-time PCR machines with a suitable fluorescent detector are used to compare melting curve shifts in Tm. Bioinformatics' tools have been developed to aid in the interpretation of Tm curves (Rosa et al. 2015) (Table 1). The ThermoFluor method has the advantage of requiring low amounts of protein: ~1.2 μg of protein at least 75% pure per well of a 96-well plate (reference weight 30 kDa) (Boivin et al. 2013). This method is also useful to evaluate the effect of fusion tags on protein stability. For example, although small, the His-tag decreased the thermal stability of a viral protein by 4.1 °C (Boivin et al. 2013).

Nevertheless, it has also some disadvantages (Lebendiker and Danieli 2014): it cannot provide information regarding the oligomeric state of the protein, it cannot be used in the presence of detergents, and the presence of intrinsic fluorescent aggregates makes it difficult to interpret. Thus, protein homogeneity must be evaluated after buffer optimization (see "Protein homogeneity" section) (Fig. 1).

Protein quantification

Accurate determination of the concentration of a purified protein is particularly important for a reliable characterization by structural and functional methods. However, the more accurate methods are not commonly accessible (e.g. quantitative amino acid analysis (AAA), Rutherfurd and Dunn 2011). Nonetheless, there are a number of easy and reliable techniques for measuring protein concentration. The classical dye-based methods make use of the binding or formation of a chromophore in the presence of the soluble protein and subsequent measurement of the chromophore absorbance (e.g. bicinchoninic acid (BCA), Bradford, Lowry). Nevertheless, these assays have advantages and disadvantages regarding accuracy, robustness or compatibility with various buffer components (Olson 2016). There are however commercially available reagents with improved performance, and optimized for fast and convenient protein determination. It is noteworthy that quantification by these methods is based on calibration with a standard protein (e.g. BSA), and dye may interact differently with different proteins, which can be a source of error (Olson 2016). It would be preferable to use as standard the protein under study, if commercially available, or a close-related protein.

Measuring the absorption at 280 nm (A_{280}) is the simplest method for determining protein concentration of pure proteins in clear solutions. This method is based on UV light absorption by aromatic residues, mainly tryptophans and tyrosines, the two main light-absorbing amino acids. Its main advantage over the dye-based methods is that it is independent of a reference protein, and so, if correctly executed, it may have a superior accuracy. It has also some disadvantages, such as the interference of nucleic acids, chromophores and detergents. For example, imidazole, typically used in His-tag purification for protein elution, absorbs UV radiation at 280 nm, and so must be removed before the measurement. Moreover, the accuracy of the method strongly depends on the correct determination of the extinction coefficient (ε) (Grimsley and Pace 2004). If this parameter has already been experimentally determined, the concentration of the protein (c) can be calculated using the Lambert-Beer law ($A_{280} = \varepsilon \times lc$, where l is the path length of the spectrometer) (Grimsley and Pace 2004; Olson 2016). If not available, ε can be estimated from the primary sequence of the protein using online tools, such as Expasy's ProtParam tool (Table 1). However, in this case, the structure is not taken into account, which can affect the

extinction coefficient. Moreover, if the protein has no tryptophans in its composition, the tool notifies for at least a 10% error in the given parameter. The tool also provides a calculated A_{280} value for a concentration of 1 mg/ml, which can be used to interpolate protein concentration. This value must be corrected for the differences between native and denatured protein. This can be done by measuring the A_{280} of the protein, in the linear range of response, in the absence (native) and presence (denatured) of 6 M guanidine (i.e. by diluting the sample four times only in buffer and in 8 M GdmCl prepared in buffer, respectively) (Kelly et al. 2005). The ratio of the A_{280} values of native and denatured proteins is normally within 0.9 and 1.1 and can be used to correct the calculated A_{280} to give a more accurate value for the native protein (Kelly et al. 2005). Alternatively, protein concentration can be determined from Beer's law using predicted ε (the value given for *all Cys residues reduced*, if protein contains cysteines in its composition) and A_{280} determined experimentally for denatured protein sample. In the absence of tryptophans and tyrosines in the protein composition, a calibration curve of A_{280} (or another wavelength) versus protein concentration determined by, e.g. quantitative AAA can be constructed and used for determination of protein concentration of subsequent samples by UV absorbance spectroscopy (Raynal et al. 2014).

Protein storage

The conditioning of a protein is as important as its production and purification. Protein properties must be preserved in storage. Storage conditions depend on the protein, and thus, there are no standard protocols. There are, however, some rules that can be followed. The first one is to avoid storage of the protein at pH values close to its isoelectric point (Table 1); otherwise, protein will precipitate. Drastic changes in pH and temperature, and overly high concentrations, should also be avoided as they may promote protein denaturation/precipitation (Jamrichová et al. 2017). Buffers reported as a source of sample heterogeneity, such as Tris, should not be used (Boivin et al. 2013). Moreover, the addition of components that may interfere with the protein application should not be an option, as this will require a further desalting or dialysis step. The less the protein is manipulated, the best for its stability. Maintaining the protein on ice (or at 4 °C) and using it in the following hours would be the better procedure; however, this is not always possible. In fact, most proteins can be kept stable at 4 °C from some days to weeks. In this last case, samples must be filter-sterilized (through a 0.22-μm filter with low binding capacity), or supplemented with antibacterial and antimycotic agents (e.g. 0.1% sodium azide), to avoid microbial contamination (Jamrichová et al. 2017). The most suitable storage temperature can be determined experimentally by monitoring the stability of small aliquots of the protein over time at some relevant temperatures (i.e. room temperature,

4 °C, or lower temperatures as − 20 °C) using, e.g. DLS and a functional assay (Raynal et al. 2014). For storage of proteins by longer periods, different methods can be tested, such as freezing, salt precipitation or lyophilization (Carpenter et al. 2002; Simpson 2010), but their effects on protein properties should be as well evaluated. It should be noted that, although a rapid freezing of small aliquots at − 20 °C is preferred over a slow freezing, this method requires case-to-case validation, as freezing/thawing can induce protein denaturation, aggregation and precipitation (Cao et al. 2003). Several additives can be included to enhance protein stability, such as cryoprotectants (e.g. glycerol, up to 40% (*w/v*)), reducing agents (e.g. 1 mM DTT), protein-specific ligands (Lebendiker and Danieli 2014) or serum albumins (e.g. BSA, 10 mg/ml) (Jamrichová et al. 2017). At the end, the optimal storage conditions can differ significantly from the application conditions, and thus, a buffer exchange step may be necessary, followed by verification of the homogeneity, functionality and concentration of the protein.

Final remarks and future perspectives

The greatest challenge in a RPP process is to obtain soluble, homogeneous, pure protein samples, natively active at well-known concentrations in suitable buffers for the aimed applications. The increasing awareness of the importance of this goal for successful recombinant proteins' use leads to significant advances in this field. Novel affinity purification tags from natural origin, genetically modified or synthetic, as well as tag combinations, have revolutionized the access of high soluble yields of difficult-to-express proteins (Yadav et al. 2016). Moreover, fusion tag technology has been fundamental for the improvement of the throughput capacity of RPP (Jia and Jeon 2016; Konczal and Gray 2017), allowing this to keep pace with the rapidly growing high-throughput omics technologies (Sequeira et al. 2017). Tag choice and the design of the purification procedure are the first-line determinants of protein quality and thus must be carefully addressed. Classical structural characterization techniques are conducted to ascertain protein quality in terms of purity (SDS-PAGE), homogeneity (DLS, SEC) and structural conformity (CD). The rising demand for multiple high-quality protein samples has led to important progresses in the high-throughput of classical techniques for assessing and optimizing protein homogeneity, solubility and stability, such as SEC (Brusotti et al. 2017) or DSF (Tileva et al. 2017). Besides, structural characterization techniques together with the development of bioinformatic tools lead to a more comprehensive knowledge on protein structure, function and environmental condition interrelation, which will pave the way for the rational design of an effective purification strategy that results in high-quality functional recombinant protein.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Agostini F, Cirillo D, Livi CM, Delli Ponti R, Tartaglia GG (2014) ccSOL omics: a webserver for solubility prediction of endogenous and heterologous expression in *Escherichia coli*. Bioinformatics 30(20): 2975–2977. https://doi.org/10.1093/bioinformatics/btu420

Aguiar TQ, Oliveira C, Domingues L (2017) Synthesis of fusion genes for cloning by megaprimer-based PCR. Methods Mol Biol 1620: 101–112. https://doi.org/10.1007/978-1-4939-7060-5_6

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(Web Server issue):W252–W258. https://doi.org/10.1093/nar/gku340

Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif 91(2):192–206. https://doi.org/10.1016/j.pep.2013.08.002

Brusotti G, Calleri E, Colombo R, Massolini G, Rinaldi F, Temporini C (2017) Advances on size exclusion chromatography and applications on the analysis of protein biopharmaceuticals and protein aggregates: a mini review. Chromatographia, In press. https://doi.org/10.1007/s10337-017-3380-5

Cantu-Bustos JE, Vargas-Cortez T, Morones-Ramirez JR, Balderas-Renteria I, Galbraith DW, McEvoy MM, Zarate X (2016) Expression and purification of recombinant proteins in *Escherichia coli* tagged with the metal-binding protein CusF. Protein Expr Purif 121:61–65. https://doi.org/10.1016/j.pep.2016.01.007

Cao E, Chen Y, Cui Z, Foster PR (2003) Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. Biotechnol Bioeng 82(6):684–690. https://doi.org/10.1002/bit.10612

Carpenter JF, Chang BS, Garzon-Rodriguez W, Randolph TW (2002) Rational design of stable lyophilized protein formulations: theory and practice. Pharm Biotechnol 13:109–133. https://doi.org/10.1007/978-1-4615-0557-0_5

Chen X, Zaro JL, Shen WC (2013) Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 65(10):1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

Chen Z, Li Y, Yuan Q (2015) Study the effect of His-tag on chondroitinase ABC I based on characterization of enzyme. Int J Biol Macromol 78:96–101. https://doi.org/10.1016/j.ijbiomac.2015.03.068

Chevallet M, Luche S, Rabilloud T (2006) Silver staining of proteins in polyacrylamide gels. Nat Protoc 1(4):1852–1858. https://doi.org/10.1038/nprot.2006.288

Costa S, Almeida A, Castro A, Domingues L (2014) Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. Front Microbiol 5:63. https://doi.org/10.3389/fmicb.2014.00063

Costa SJ, Almeida A, Castro A, Domingues L, Besir H (2013a) The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. Appl Microbiol Biotechnol 97(15):6779–6791. https://doi.org/10.1007/s00253-012-4559-1

Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L (2013b) The Fh8 tag: a fusion partner for simple and cost-effective protein purification in *Escherichia coli*. Protein Expr Purif 92(2): 163–170. https://doi.org/10.1016/j.pep.2013.09.013

De Jong CA, Risley J, Lee AK, Zhao SS, Chen DD (2016) Separation of recombinant therapeutic proteins using capillary gel electrophoresis and capillary isoelectric focusing. Methods Mol Biol 1466:137–149. https://doi.org/10.1007/978-1-4939-4014-1_11

Dullah EC, Ongkudon CM (2017) Current trends in endotoxin detection and analysis of endotoxin-protein interactions. Crit Rev Biotechnol 37(2):251–261. https://doi.org/10.3109/07388551.2016.1141393

England CG, Luo H, Cai W (2015) HaloTag technology: a versatile platform for biomedical applications. Bioconjug Chem 26(6):975–986. https://doi.org/10.1021/acs.bioconjchem.5b00191

Fekete S, Beck A, Veuthey JL, Guillarme D (2014) Theory and practice of size exclusion chromatography for the analysis of protein aggregates. J Pharm Biomed Anal 101:161–173. https://doi.org/10.1016/j.jpba.2014.04.011

Francis DM, Page R (2010) Strategies to optimize protein expression in *E. coli*. Curr Protoc Protein Sci Chapter 5:Unit 5 24 1–29. doi: https://doi.org/10.1002/0471140864.ps0524s61

Gieseler G, Pepelanova I, Stuckenberg L, Villain L, Nolle V, Odenthal U, Beutel S, Rinas U, Scheper T (2017) Purification of bone morphogenetic protein-2 from refolding mixtures using mixed-mode membrane chromatography. Appl Microbiol Biotechnol 101(1):123–130. https://doi.org/10.1007/s00253-016-7784-1

Glasel JA (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. BioTechniques 18(1):62–63

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 HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schutz A, Heinemann U, Yokoyama S, Bussow K, Gunsalus KC, [Structural Genomics C, China Structural Genomics C, Northeast Structural Genomics C] (2008) Protein production and purification. Nat Methods 5(2):135–146. https://doi.org/10.1038/nmeth.f.202

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1(6):2876–2890. https://doi.org/10.1038/nprot.2006.202

Grimsley GR, Pace CN (2004) Spectrophotometric determination of protein concentration. Curr Protoc Protein Sci Chapter 3:Unit 3 1. doi: https://doi.org/10.1002/0471140864.ps0301s33

Guerreiro CI, Fontes CM, Gama M, Domingues L (2008) *Escherichia coli* expression and purification of four antimicrobial peptides fused

to a family 3 carbohydrate-binding module (CBM) from *Clostridium thermocellum*. Protein Expr Purif 59(1):161–168. https://doi.org/10.1016/j.pep.2008.01.018

He F, Becker GW, Litowski JR, Narhi LO, Brems DN, Razinkov VI (2010) High-throughput dynamic light scattering method for measuring viscosity of concentrated protein solutions. Anal Biochem 399(1):141–143. https://doi.org/10.1016/j.ab.2009.12.003

Healey RD, Lebhar H, Hornung S, Thordarson P, Marquis CP (2017) An improved process for the production of highly purified recombinant thaumatin tagged-variants. Food Chem 237:825–832. https://doi.org/10.1016/j.foodchem.2017.06.018

Hebditch M, Alejandro Carballo-Amador M, Charonis S, Curtis R, Warwicker J (2017) Protein-Sol: a web tool for predicting protein solubility from sequence. Bioinformatics. https://doi.org/10.1093/bioinformatics/btx345

Hollingshead S, Lin CY, Liu JC (2017) Designing smart materials with recombinant proteins. Macromol Biosci 17(7). https://doi.org/10.1002/mabi.201600554

Jamrichová D, Tišáková L, Jarábková V, Godány A (2017) How to approach heterogeneous protein expression for biotechnological use: an overview. Nova Biotechnol Chim 16(1):1–11. https://doi.org/10.1515/nbec-2017-0001

Jia B, Jeon CO (2016) High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. Open biology 6(8). https://doi.org/10.1098/rsob.160196

Jozala AF, Geraldes DC, Tundisi LL, Feitosa VA, Breyer CA, Cardoso SL, Mazzola PG, Oliveira-Nascimento L, Rangel-Yagui CO, Magalhaes PO, Oliveira MA, Pessoa A Jr (2016) Biopharmaceuticals from microorganisms: from production to purification. Braz J Microbiol 47(Suppl 1):51–63. https://doi.org/10.1016/j.bjm.2016.10.007

Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751(2):119–139. https://doi.org/10.1016/j.bbapap.2005.06.005

Khan E, Kumar A (2017) Emerging methods for structural analysis of protein aggregation. Protein Pept Lett. https://doi.org/10.2174/0929866524666170206123150

Konczal J, Gray CH (2017) Streamlining workflow and automation to accelerate laboratory scale protein production. Protein Expr Purif 133:160–169. https://doi.org/10.1016/j.pep.2017.03.016

Lebendiker M, Danieli T (2014) Production of prone-to-aggregate proteins. FEBS Lett 588(2):236–246. https://doi.org/10.1016/j.febslet.2013.10.044

Lebendiker M, Danieli T, de Marco A (2014) The Trip Adviser guide to the protein science world: a proposal to improve the awareness concerning the quality of recombinant proteins. BMC Res Notes 7:585. https://doi.org/10.1186/1756-0500-7-585

Li XJ, Liu JL, Gao DS, Wan WY, Yang X, Li YT, Chang HT, Chen L, Wang CQ, Zhao J (2016) Single-step affinity and cost-effective purification of recombinant proteins using the Sepharose-binding lectin-tag from the mushroom *Laetiporus sulphureus* as fusion partner. Protein Expr Purif 119:51–56. https://doi.org/10.1016/j.pep.2015.11.004

Li YF (2011) The tandem affinity purification technology: an overview. Biotechnol Lett 33(8):1487–1499. https://doi.org/10.1007/s10529-011-0592-x

Lin Z, Zhao Q, Xing L, Zhou B, Wang X (2015) Aggregating tags for column-free protein purification. Biotechnol J 10(12):1877–1886. https://doi.org/10.1002/biot.201500299

Magalhães F, Aguiar TQ, Oliveira C, Domingues L (2014) High-level expression of *Aspergillus niger* β-galactosidase in *Ashbya gossypii*. Biotechnol Prog 30(2):261–268. https://doi.org/10.1002/btpr.1844

Manta B, Obal G, Ricciardi A, Pritsch O, Denicola A (2011) Tools to evaluate the conformation of protein products. Biotechnol J 6(6): 731–741. https://doi.org/10.1002/biot.201100107

Minton AP (2016) Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. Anal Biochem 501: 4–22. https://doi.org/10.1016/j.ab.2016.02.007

Morris J, Jayanthi S, Langston R, Daily A, Kight A, McNabb DS, Henry R, Kumar TKS (2016) Heparin-binding peptide as a novel affinity tag for purification of recombinant proteins. Protein Expr Purif 126: 93–103. https://doi.org/10.1016/j.pep.2016.05.013

Murphy RM, Roberts CJ (2013) Protein misfolding and aggregation research: some thoughts on improving quality and utility. Biotechnol Prog 29(5):1109–1115. https://doi.org/10.1002/btpr.1812

Nguyen MT, Krupa M, Koo BK, Song JA, Vu TT, Do BH, Nguyen AN, Seo T, Yoo J, Jeong B, Jin J, Lee KJ, Oh HB, Choe H (2016) Prokaryotic soluble overexpression and purification of human VEGF165 by fusion to a maltose binding protein tag. PLoS One 11(5):e0156296. https://doi.org/10.1371/journal.pone.0156296

Noirclerc-Savoye M, Flayhan A, Pereira C, Gallet B, Gans P, Ebel C, Breyton C (2015) Tail proteins of phage T5: investigation of the effect of the His6-tag position, from expression to crystallisation. Protein Expr Purif 109:70–78. https://doi.org/10.1016/j.pep.2015.02.003

Olson BJ (2016) Assays for determination of protein concentration. Curr Protoc Pharmacol 73:A 3A 1-A 3A 32. doi:https://doi.org/10.1002/cpph.3

Oliveira C, Carvalho V, Domingues L, Gama FM (2015) Recombinant CBM-fusion technology—applications overview. Biotechnol Adv 33(3–4):358–369. https://doi.org/10.1016/j.biotechadv.2015.02.006

Oliveira C, Aguiar TQ, Domingues L (2017) 4—Principles of genetic engineering. In: Pandey A, Teixeira JA (eds) Current developments in biotechnology and bioengineering: foundations of biotechnology and bioengineering, 1st edn. Elsevier, Oxford, pp 81–127 ISBN 9780444636683

Oliveira C, Felix W, Moreira RA, Teixeira JA, Domingues L (2008) Expression of frutalin, an alpha-D-galactose-binding jacalin-related lectin, in the yeast *Pichia pastoris*. Protein Expr Purif 60(2):188–193. https://doi.org/10.1016/j.pep.2008.04.008

Oliveira C, Teixeira JA, Domingues L (2014a) Recombinant production of plant lectins in microbial systems for biomedical application—the frutalin case study. Front Plant Sci 5:390. https://doi.org/10.3389/fpls.2014.00390

Oliveira H, Thiagarajan V, Walmagh M, Sillankorva S, Lavigne R, Neves-Petersen MT, Kluskens LD, Azeredo J (2014b) A thermostable *Salmonella* phage endolysin, Lys68, with broad bactericidal properties against gram-negative pathogens in presence of weak acids. PLoS One 9(10):e108376. https://doi.org/10.1371/journal.pone.0108376

Pallares I, Ventura S (2017) Advances in the prediction of protein aggregation propensity. Curr Med Chem. https://doi.org/10.2174/0929867324666170705121754

Rabilloud T (2012) Silver staining of 2D electrophoresis gels. Methods Mol Biol 893:61–73. https://doi.org/10.1007/978-1-61779-885-6_5

Ramos R, Domingues L, Gama M (2010) *Escherichia coli* expression and purification of LL37 fused to a family III carbohydrate-binding module from *Clostridium thermocellum*. Protein Expr Purif 71(1): 1–7. https://doi.org/10.1016/j.pep.2009.10.016

Ramos R, Moreira S, Rodrigues A, Gama M, Domingues L (2013) Recombinant expression and purification of the antimicrobial peptide magainin-2. Biotechnol Prog 29(1):17–22. https://doi.org/10.1002/btpr.1650

Raynal B, Lenormand P, Baron B, Hoos S, England P (2014) Quality assessment and optimization of purified protein samples: why and how? Microb Cell Factories 13:180. https://doi.org/10.1186/s12934-014-0180-6

Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS (2013) Optimization of protein buffer cocktails using Thermofluor. Acta Crystallogr Sect F Struct Biol Cryst Commun 69(Pt 2):209–214. https://doi.org/10.1107/S1744309112051858

Ribeiro O, Wiebe M, Ilmen M, Domingues L, Penttila M (2010) Expression of *Trichoderma reesei* cellulases CBHI and EGI in *Ashbya gossypii*. Appl Microbiol Biotechnol 87(4):1437–1446. https://doi.org/10.1007/s00253-010-2610-7

Rosa N, Ristic M, Seabrook SA, Lovell D, Lucent D, Newman J (2015) Meltdown: a tool to help in the interpretation of thermal melt curves acquired by differential scanning fluorimetry. J Biomol Screen 20(7):898–905. https://doi.org/10.1177/1087057115584059

Rutherfurd SM, Dunn BM (2011) Quantitative amino acid analysis. Curr Protoc Protein Sci Chapter 3:Unit 3 2. doi:https://doi.org/10.1002/0471140864.ps0302s63

Saccardo P, Corchero JL, Ferrer-Miralles N (2016) Tools to cope with difficult-to-express proteins. Appl Microbiol Biotechnol 100(10): 4347–4355. https://doi.org/10.1007/s00253-016-7514-8

Saraswat M, Musante L, Ravida A, Shortt B, Byrne B, Holthofer H (2013) Preparative purification of recombinant proteins: current status and future trends. Biomed Res Int 2013:312709. https://doi.org/10.1155/2013/312709

Senisterra G, Chau I, Vedadi M (2012) Thermal denaturation assays in chemical biology. Assay Drug Dev Technol 10(2):128–136. https://doi.org/10.1089/adt.2011.0390

Sequeira AF, Bras JLA, Fernandes VO, Guerreiro C, Vincentelli R, Fontes C (2017) A novel platform for high-throughput gene synthesis to maximize recombinant expression in *Escherichia coli*. Methods Mol Biol 1620:113–128. https://doi.org/10.1007/978-1-4939-7060-5_7

Shi C, Meng Q, Wood DW (2013) A dual ELP-tagged split intein system for non-chromatographic recombinant protein purification. Appl Microbiol Biotechnol 97(2):829–835. https://doi.org/10.1007/s00253-012-4601-3

Shi S, Chen H, Jiang H, Xie Y, Zhang L, Li N, Zhu C, Chen J, Luo H, Wang J, Feng L, Lu H, Zhu J (2017) A novel self-cleavable tag Zbasic-I-CM and its application in the soluble expression of recombinant human interleukin-15 in *Escherichia coli*. Appl Microbiol Biotechnol 101(3):1133–1142. https://doi.org/10.1007/s00253-016-7848-2

Simpson RJ (2010) Stabilization of proteins for storage. Cold Spring Harb Protoc 2010(5):pdb top79. doi:https://doi.org/10.1101/pdb.top79

Stetefeld J, McKenna SA, Patel TR (2016) Dynamic light scattering: a practical guide and applications in biomedical sciences. Biophys Rev 8(4):409–427. https://doi.org/10.1007/s12551-016-0218-6

Tileva M, Krachmarova E, Taneva SG, Todinova S, Maskos K, Ivanov I, Nacheva G (2017) Buffer and additive thermofluor screening of wild type human interferon gamma and mutant proteins. Thermochim Acta 654:1–7. https://doi.org/10.1016/j.tca.2017.05.003

Vargas-Cortez T, Morones-Ramirez JR, Balderas-Renteria I, Zarate X (2016) Expression and purification of recombinant proteins in *Escherichia coli* tagged with a small metal-binding protein from *Nitrosomonas europaea*. Protein Expr Purif 118:49–54. https://doi.org/10.1016/j.pep.2015.10.009

Vedadi M, Arrowsmith CH, Allali-Hassani A, Senisterra G, Wasney GA (2010) Biophysical characterization of recombinant proteins: a key to higher structural genomics success. J Struct Biol 172(1):107–119. https://doi.org/10.1016/j.jsb.2010.05.005

Wang K, Sun D-W, Pu H, Wei Q (2017) Principles and applications of spectroscopic techniques for evaluating food protein conformational changes: a review. Trends Food Sci Technol 67:207–219. https://doi.org/10.1016/j.tifs.2017.06.015

Wang W, Nema S, Teagarden D (2010) Protein aggregation—pathways and influencing factors. Int J Pharm 390(2):89–99. https://doi.org/10.1016/j.ijpharm.2010.02.025

Yadav DK, Yadav N, Yadav S, Haque S, Tuteja N (2016) An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. Arch Biochem Biophys 612:57–77. https://doi.org/10.1016/j.abb.2016.10.012

Yang Y, Niroula A, Shen B, Vihinen M (2016) PON-Sol: prediction of effects of amino acid substitutions on protein solubility. Bioinformatics 32(13):2032–2034. https://doi.org/10.1093/bioinformatics/btw066

Zhang G, Annan RS, Carr SA, Neubert TA (2014) Overview of peptide and protein analysis by mass spectrometry. Curr Protoc Mol Biol 108:10 21 1–30. doi:https://doi.org/10.1002/0471142727.mb1021s108

Zhang K, Liu X (2016) Mixed-mode chromatography in pharmaceutical and biopharmaceutical applications. J Pharm Biomed Anal 128:73–88. https://doi.org/10.1016/j.jpba.2016.05.007

Zou L, Su W, Wang M, Huang W, Zhao H, Zhang E, Jin J, Xu H, Xiao F (2017) Characterization of a functional recombinant human creatine kinase-MB isoenzyme prepared by tandem affinity purification from *Escherichia coli*. Appl Microbiol Biotechnol 101(14):5639–5644. https://doi.org/10.1007/s00253-017-8286-5

Zvonova EA, Ershov AV, Ershova OA, Sudomoina MA, Degterev MB, Poroshin GN, Eremeev AV, Karpov AP, Vishnevsky AY, Goldenkova-Pavlova IV, Petrov AV, Ruchko SV, Shuster AM (2017) PASylation technology improves recombinant interferon-beta1b solubility, stability, and biological activity. Appl Microbiol Biotechnol 101(5):1975–1987. https://doi.org/10.1007/s00253-016-7944-3