

A Screening Methodology for Purifying Proteins with Aggregation Problems

Mario Lebendiker, Michal Maes, and Assaf Friedler

Abstract

Many proteins are prone to aggregate or insoluble for different reasons. This poses an extraordinary challenge at the expression level, but even more during downstream purification processes. Here we describe a strategy that we developed for purifying prone-to-aggregate proteins. Our methodology can be easily implemented in small laboratories without the need for automated, expensive platforms. This procedure is especially suitable for intrinsically disordered proteins (IDPs) and for proteins with intrinsically disordered regions (IDRs). Such proteins are likely to aggregate due to their lack of tertiary structure and their extended and flexible conformations. Similar methodologies can be applied to other proteins with comparable tendency to aggregate during the expression or purification steps.

In this chapter, we will mainly focus on protein solubility and stability issues during purification and storage, on factors that can prevent aggregation or maintain solubility, and on the importance of the early elimination of aggregates during protein purification.

Key words Protein aggregation, Insoluble proteins, Intrinsically disordered proteins, Protein storage, Protein concentration, Stabilizers, Aggregation suppressors, Chaotropes, Kosmotropes, Buffer conditions, Aggregation analysis

1 Introduction

1.1 *Insoluble Proteins, Instability, and Aggregation*

Stability is an extremely important issue in protein production, due to the fact that once destabilized, proteins are susceptible to chemical and physical alteration that lead to loss of activity. Chemical alteration as protein cleavage or related to covalent bond modifications like oxidation and disulfide bond shuffling. Physical changes include protein unfolding, undesirable binding to surfaces, and aggregation [1]. These undesirable changes can be reversible or irreversible. They can produce aggregates that range in size from soluble aggregates, only detectable by size exclusion chromatography (SEC), to particles that may contain trillions (or more) of monomer units visible by the eye [2]. There is a great concern about the presence of aggregates in therapeutic proteins because of

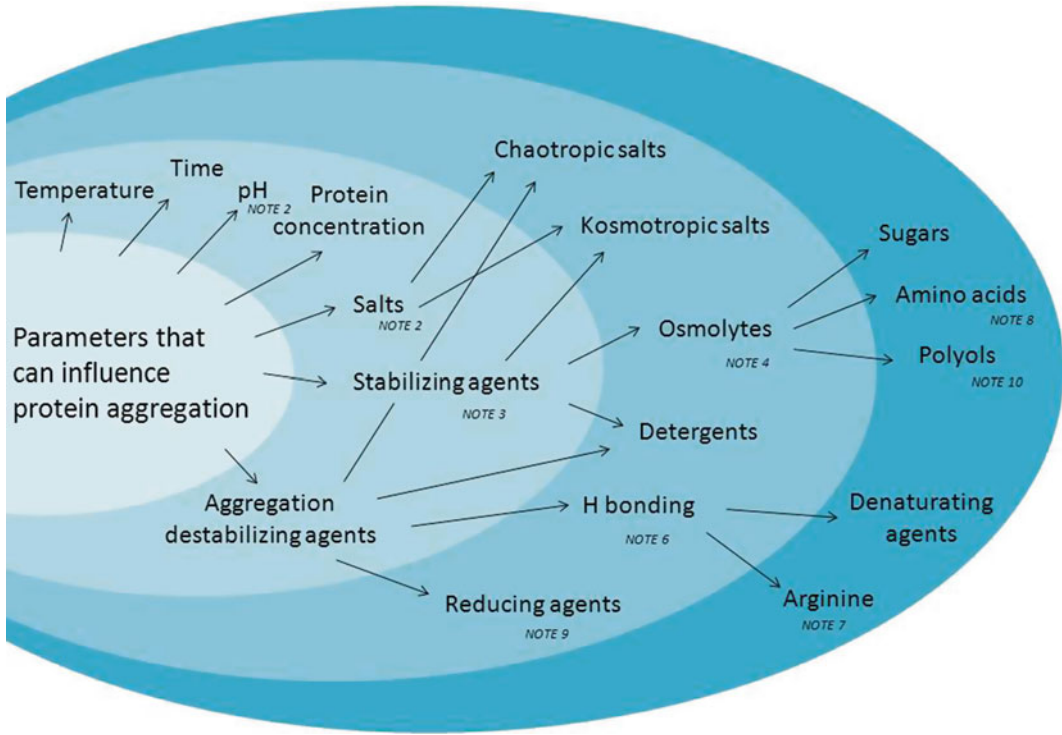


Fig. 1 Critical issues to be considered in order to prevent aggregation during protein purification

their unpredictable ability to give rise to adverse toxicological and immunological responses, which in extreme cases can be life-threatening [2]. As the number of therapeutic proteins increases, finding ways to understand and prevent this problem continues to gain importance. The same issues of protein instability and aggregation cause many problems in basic as well as in applied research: protein production yields are decreased, aggregated proteins are unable to crystallize, their specific activity is highly affected, and the credibility of the results using aggregated proteins in all kinds of experiments is questionable.

Aggregation is an undesired interaction between protein monomers. This process can be influenced by temperature, protein concentration, buffer conditions, etc. (Fig. 1). There is an extended lag phase before large aggregates appear and accumulate in an abrupt way [3].

Protein aggregates may be classified in numerous ways, including soluble/insoluble, covalent/non-covalent, reversible/irreversible, native/denatured, or by size, conformation, and morphology [4, 5]. Some efforts are made for nomenclature standardization and classification [4]. Five major mechanisms of aggregation have been proposed: concentration-induced aggregation, aggregation induced by conformational changes, aggregation induced by

chemical reactions, nucleation-dependent aggregation, and surface-induced aggregation [3, 6]. A fundamental understanding of the mechanism of aggregation is not only valuable for identifying the cause of the problem but is also helpful for developing methods to suppress aggregation [3, 6].

Insolubility of recombinant proteins may be encountered already at the expression level. Several solutions can overcome this problem. These include screening of different bacterial strains, decreasing culture temperatures, different culture mediums, different fusion protein constructs such as maltose binding protein (*see* Chapter 2), alternative expression systems such as cell-free expression (*see* Chapter 6) or baculovirus (*see* Chapter 9), using constructs with either amino or carboxyl-terminal deletions, expression of homologs of a protein of interest, removing flexible loops or residues that affect solubility, and refolding of denatured proteins [7]. As detailed, many of these approaches will be extensively discussed in other chapters.

1.2 Factors that Influence Protein Aggregation and Insolubility During Expression and Purification

Although protein solubility during expression is an essential prerequisite before purification, this does not prevent aggregation problems from arising at later stages of the protein production process (*see* Note 1). In this chapter, we will focus mainly on protein stability issues that must be considered from the very early purification steps until storage. Some general issues that can lead to denaturation and aggregation and should be considered are purification time and temperature, protein concentration at each step, and prevention of mechanical or nonmechanical stresses (freezing, exposure to air, interactions with metal surfaces, etc.). Other factors that can influence aggregation are pH or ionic strength. In addition, the protein environment can be affected by cosolutes such as chaotropes and kosmotropes (*see* Note 2), osmolytes and ligands, protein–protein interaction inhibitors, reducing agents, surfactants, and non-denaturative detergents. All of these can be divided into two main categories: factors that stabilize proteins and factors that inhibit aggregation or inhibit protein–protein interactions (Fig. 1, *see* Note 3).

A change in solution conditions such as a decrease in protein concentration or changes in pH or salt concentration can dissociate the aggregates in some cases. This is especially true for aggregates where the molecules are held together by relatively weak, non-covalent interactions. However, such changes rarely affect other types of aggregation. Such pH- or salt-dependent reversibility is indicative of equilibrium between the monomer and high-order forms [5].

Protein stabilizers are additives that inhibit aggregation by stabilizing the native structure of the protein [8]. There is correlation between additives that stabilize proteins against thermal stress in cells and additives that stabilize proteins during isolation and

storage. These stabilizing cosolutes are also termed osmolytes, since they are utilized in nature to increase the osmotic pressure of the cellular environment and are compatible with the macromolecular function and cell viability (*see Note 4*) [8]. Examples for such osmolytes are trehalose and trimethylamine *N*-oxide (TMAO), both used for protein refolding [9]. Other examples include sucrose, glycerol, sorbitol, mannitol, glycine betaine (betaine) [10], and proline [11]. Polyethylene glycol (PEG) (*see Note 5*) and kosmotropic salts as magnesium or ammonium sulfate [8, 12, 13] and potassium citrate [7] also act as protein stabilizers. Alcohols such as ethanol can be used to stabilize folding intermediates by weakening hydrophobic interactions that facilitate aggregation [12] (unpublished data).

Aggregation suppressors can work in several ways. The H-bonding agents, like urea or guanidine HCl (GdnHCl), work as chaotropic agents at low concentration (0.5–2 M). They decrease the net hydrophobic effect of prone-to-aggregate hydrophobic regions in proteins by disordering the water molecules adjacent to the protein surface (*see Note 6*). The way L-arginine hydrochloride (L-ArgHCl) protects proteins from aggregation is more complicated. It can act as an H-bonding agent like urea or GdnHCl, but it has certain kosmotropic properties, allowing it to interact with aromatic side chains of the protein (*see Note 7*) [14]. Other amino acids such as proline, histidine, and beta-alanine, as well as the naturally occurring polyamines putrescine, spermidine, and spermine, were also reported as aggregation suppressors [15].

Aggregation can be induced by chemical modifications such as incorrect disulfide bond or arrangement or the formation of bi-tyrosine (*see Note 8*). The presence of weak reducing agents and oxidants can reverse this problem or lead to changes in protein conformation that may alter the function of the protein. Reducing agents can break disulfide bonds and lead to dissociation of parts of the protein chain(s) that are normally associated. Oxidants can cause the formation of disulfide bonds and consequent association of parts of the protein chain that are normally not associated (*see Note 9*).

Surfactants are used in biotechnology to stabilize therapeutic proteins, suppress aggregation, and assist in protein refolding. They can prevent protein adsorption on surfaces, which would result in loss of activity and/or surface-induced aggregation. Surfactants can also bind hydrophobic regions in proteins and thus prevent aggregation [6]. Some widely used surfactants are polysorbate, poloxamers, and non-detergent sulfobetaines (NDSBs) (*see Note 10*) [6].

Although recommended additive concentrations are found in the literature [12, 13, 16, 17], the optimal range for each protein is highly specific, and the buffer conditions must be fine-tuned for each project (*see Note 11*). Moreover, there could be a synergistic

effect between some of these agents. This could prevent different aggregation mechanisms, for example, osmolytes, as cosolutes will favor protein structures with minimal surface area, while addition of surfactants can mask exposed hydrophobic regions [13]. Each family of additives will improve solubility of some proteins while decreasing the solubility of others. The same kosmotrope environment that stabilizes folded proteins can enhance protein–protein interactions and subsequent aggregation in partially unfolded proteins. On the other hand, chaotropic agents that destabilize aggregation of proteins in the native state can induce or enhance aggregation of partially unfolded proteins [13].

Finding the optimal buffer conditions can be performed using functional biological assays, but this is not applicable to all proteins. There could be cases where no assay is available, the assay is not reliable, or alternatively time, effort, and cost make the assay unfruitful. In any case, such assays do not provide information regarding yield, oligomeric homogeneity, and protein purity.

Several experimental methods are routinely used to determine aggregation: visual observation of turbidity, size exclusion chromatography (SEC), circular dichroism (CD), light scattering (LS), fluorescence-based thermal shift (ThermoFluor) assay, and more (*see Note 12*). No single method is optimal for all aggregates. Since there is a large number of variables to determine (different buffers, pH, additives, salt, etc.), there is a need for a progressive and rational experimental methodology that can be used to identify the optimal buffer conditions and additive concentrations to maintain protein solubility. High throughput screening (HTS) assays are not always available, so alternatives must be found (*see Note 13*). In a recent publication, Leibly et al. used a screening methodology with 144 additives, but only the classical ones gave the best results (*see Note 11*) [7]. Their findings confirm our assumption that for nonautomated laboratories, using a shorter list of additives covering most of the aggregation mechanisms can considerably reduce cost and efforts.

1.3 Our Approach for Minimizing Aggregation

We present a new approach for minimizing aggregation. Our approach is based on a hierarchical buffer selection using a small group of additives, covering different mechanism of aggregation inhibition. A similar approach has been previously reported by the Bondos' lab for pure or almost pure proteins (*see Notes 12 and 13*) [12, 13]. To maximize yield and information, we prefer to tackle the solubility issue early, starting from the cell lysate, and then continue analyzing the oligomeric state of the partially pure protein during the different purification steps until the final pure product.

Our strategy (Fig. 2) begins with a screening of solubility-promoting buffers during cell lysis, followed by a quick capture step by parallel small-scale immobilized metal chelate

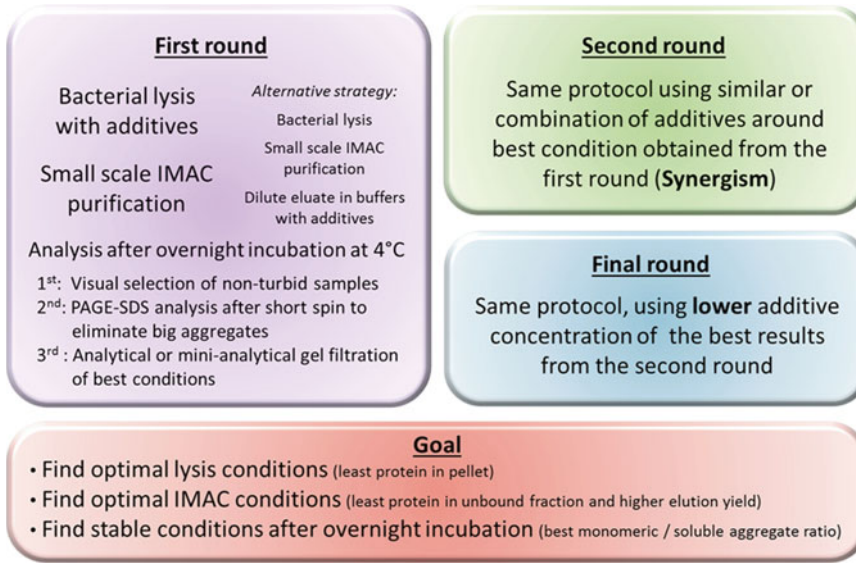


Fig. 2 Screening methodology for purifying proteins with aggregation problems

chromatography/Ni column (IMAC) purification (or any other capture method) in the presence of selected additives. Analysis is performed by SDS-PAGE of the insoluble lysis extract, unbound fraction, and eluted protein. Only the best elution conditions are further analyzed by analytical SEC, immediately after elution and after 24 h at 4 °C (time dependent aggregation), searching for the best monomer/soluble aggregate ratio (Fig. 3b). From this first screening, it is possible to estimate the influence of different additives groups on insoluble aggregated proteins, allowing better binding to the capture resin and as a result obtaining the best yield of native oligomeric conformation. This strategy not only provides maximum information on solubility issues but also improves the final output, since it rescues the protein fraction that was initially soluble in the bacteria but was then secluded as insoluble protein [7]. The following optimization rounds (Fig. 2) check if other additives from the same category may give better results, together with a combination of agents that can synergize protein solubility. In the final step, the additives concentration is optimized together with stability over time (Fig. 3b). For some projects, different buffers, pH, and additive type and concentration must be matched for each purification step. In these cases, before scaling up, the best additive(s) must be found for all intermediate steps and for storage (*see Note 14*). In some cases, the beneficial effect of the chosen additives may be maximal during cell lysis and early purification steps. This will allow drastic reduction of their concentrations at later stages [7].

Since aggregation is a nucleation-growth process, the presence of soluble aggregates during bacterial lysis can accelerate the

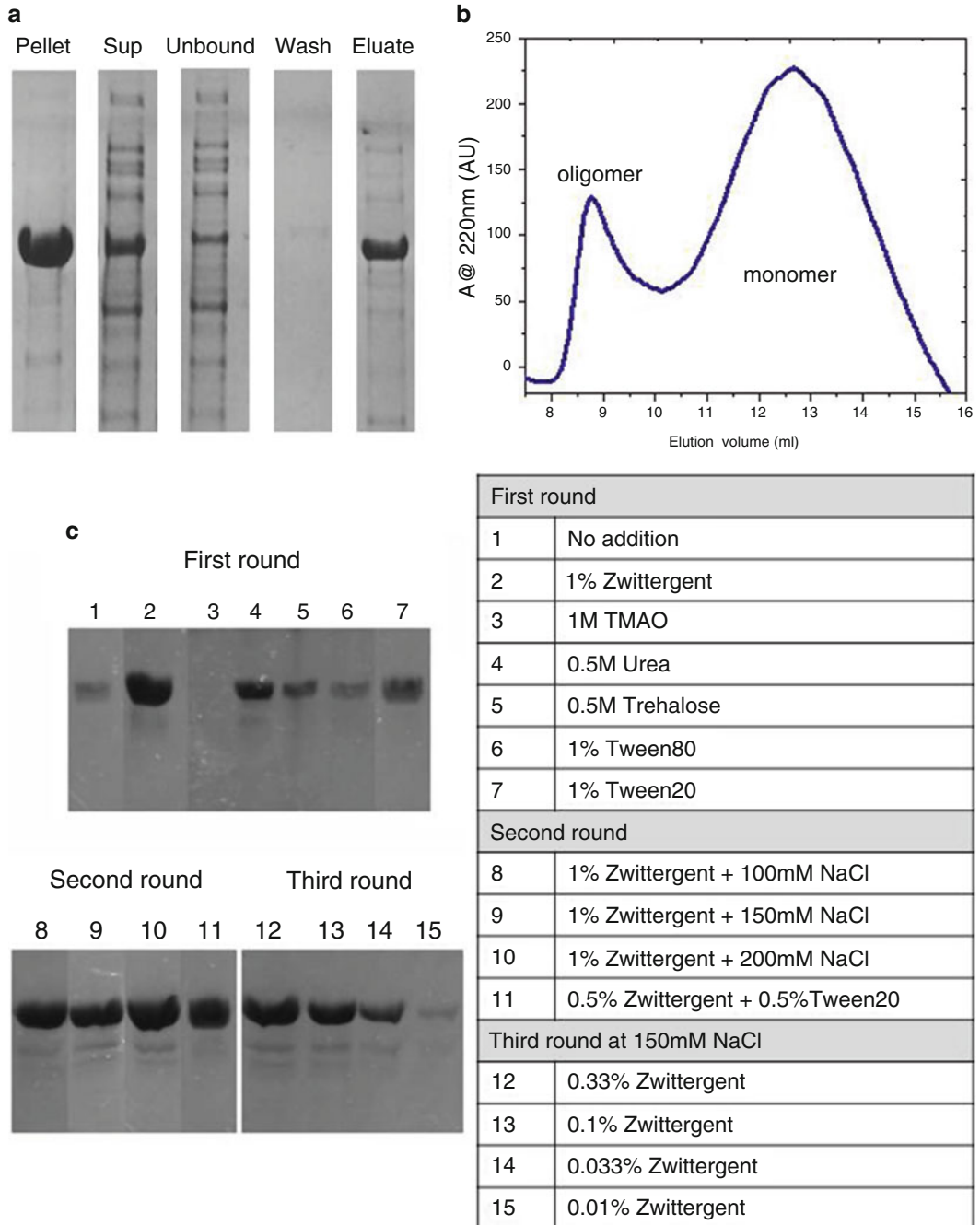


Fig. 3 Example of screening methodology. **(a)** small-scale IMAC purification; **(b)** spectrum of analytical gel filtration – the protein appears in two distinct peaks, oligomer and monomer; **(c)** three rounds of optimization for additives, according to the table (Data contributed by Dr. Ronen Gabizon)

insolubilization process. Thus, the classical strategy used by many of the protein-producing laboratories (first capture on IMAC column, followed by protease cleavage under dialysis and negative IMAC, and final polishing by size exclusion chromatography) [18, 19] can be harmful while processing of prone-to-aggregate proteins. The risk is that the presence of soluble aggregates after the first IMAC column will trigger the insolubility of more protein molecules and decrease total yield (unpublished data, *see Note 15*). A better strategy for such proteins is to try and remove the soluble aggregates as soon as possible by performing SEC (or other chromatographic procedures, *see Note 16*) immediately after the IMAC purification, followed by tag cleavage. In extremely problematic projects, we observed that high protein concentration during cell lysis or at the top of the column during chromatographic loading could sometimes speed up the aggregation process. These problems can be overcome by higher lysis volume or batch purification (*see Note 17*) or by immediate dilution of the concentrated protein after elution (*see Note 18*).

Once the conditions that give the optimal ratio between active protein and unusable aggregates are found, they must be checked for suitability with long-term storage or certain particular experimental requirements (NMR, crystallography, etc.). In some cases, an additional screening will be required to determine the buffer conditions appropriate for storage and specific usage (unpublished data, *see Note 19*).

Finally, the importance of designing a “quick strategy of purification” must be emphasized, since process time is one of the most critical points to consider. Pure protein should be produced and stored as quickly as possible. For this reason, maximum efforts must be made to optimize and fine-tune each purification step before scale-up, guaranteeing that the whole process can be performed quickly and smoothly.

The experience accumulated in our laboratory using these approaches with many IDPs and IDRs is useful for project-oriented protein production of prone-to-aggregated proteins in academic and nonautomated laboratories (without standard HTS).

2 Materials

2.1 First Round of Buffer Additives

1. Basic lysis/wash buffer: 50 mM Tris-HCl, 500 mM NaCl with 10 % glycerol, pH 8.0, with/without β -mercaptoethanol (BME) (*see Note 9*), and different *additives*.
2. Lysis buffer: wash buffer, 1 mM PMSF, lysozyme 0.2 mg/mL, DNase 50 μ g/mL, protease inhibitor cocktail.
3. Elution buffer: wash buffer, 300 mM imidazole, and *additives*.

4. *Additives*: (a) 1 M guanidine HCl, (b) 1 M urea, (c) 0.5 % Tween 20, (d) 0.5 % *n*-tetradecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate (Zwittergent 3-14), (e) 0.5 M trehalose, (f) 500 mM L-ArgHCl (only in the elution buffer).

2.2 Cell Lysis

1. Microfluidizer (LV1, Microfluidics Corp., Newton, MA) or Sonicator (Sonics Vibra Cell VCX 750) for small scale (less than 10 mL).
2. Microfluidizer (M-110 EHIS, Microfluidics Corp., Newton, MA) for large volumes (more than 10 mL).

2.3 Small-Scale IMAC Purification

1. IMAC beads: Ni-NTA or similar beads for small-scale batch purification.
2. IMAC beads: Ni Sepharose High Performance or similar beads for large-scale column purification.

2.4 Analytical and Mini-Analytical Size Exclusion Chromatography (SEC)

1. ÄKTA explorer system (GE Healthcare).
2. Analytical SEC Superdex™ 200 or 75 HR 10/30 or Superose 12 30×1 (GE Healthcare). Use according to molecular weight of the protein. Flow: 0.7 mL/min.
3. Mini-Analytical SEC: homemade columns using Superdex™ 200, Superdex™ 75, Superose 12 resin, and Tricorn 5/200 column (~4 mL) (GE Healthcare). Flow: 0.3 mL/min.

2.5 Second Round: Buffer Optimization

pH optimization: prepare several buffers changing two variables: pH and conductivity

1. 50 mM MES, pH 6.0.
2. 50 mM phosphate buffer, pH 7.0.
3. 50 mM Tris-HCl, pH 8.5.
4. 50, 300, and 500 mM NaCl.
5. Different *additives* to each buffer (*see Note 9*).

2.6 Set Up Concentration Limit

Disposable 0.5 mL ultrafiltration devices or protein concentrators with molecular weight cutoff lower than that of the native protein.

3 Methods

3.1 First Buffer Selection: Different Types of Additives (Fig. 2)

Prepare wash, lysis, and elution buffers with the different *additives* and with or without BME. Each *additive* represents a different mechanism of protein stabilization or suppression of aggregation (*see* Subheading 2.1 and *Note 9*). Add L-ArgHCl only in the elution buffer.

3.2 Cell Lysis and Clarification

1. Grow bacterial cells and induce protein expression according to the best overexpression conditions found (temperature, time, [IPTG], induction time, etc.).
2. Harvest cells and keep aliquots of 15, 100, and 500 mL pellet cells at -80°C until further processing (100 and 500 mL aliquots will be used for future scale-up).
3. Resuspend different pellets from 15 mL cell culture in 1.5 mL lysis buffer with different additives (*see* Subheading 2.1) and lyse mechanically using a Microfluidizer at 21,000 psi at 4°C or sonication on ice for 3×10 s or more (*see* Subheading 2.2) if the cells are not completely disrupted (lysis is complete when the cloudy cell suspension becomes translucent; avoid protein denaturation by frothing and extensive sonication). Remove insoluble cell debris from the cell lysate by centrifugation at 4°C for 20 min $18,000 \times g$. Separate clear supernatant (lysate) from the pellet. Keep sample of supernatant for further analysis by SDS-PAGE or Western Blot: supernatant. Continue with supernatant (*see* Subheading 3.3).
4. Resuspend pellet (insoluble cell debris) in 1.5 mL buffer and keep sample for further analysis by SDS-PAGE or Western Blot: pellet.

3.3 Small-Scale IMAC Purification

1. Equilibration of IMAC beads: place 200 μL IMAC beads in a 2 mL plastic centrifuge tube for each condition. Wash once with 1.5 mL H_2O and twice with 1.5 mL lysis buffer (washing: mix, spin 3 min at $1,200 \times g$ discard supernatant).
2. Mix supernatant of each condition with its equivalent equilibrated resin at 4°C for 60 min.
3. Spin for 3 min at $1,200 \times g$ at 4°C . Discard supernatant and keep sample of 40 μL (unbound proteins) for PAGE-SDS or Western Blot.
4. Wash beads with 1.5 mL buffer (of each condition) at least three times: mix, spin 3 min $1,200 \times g$, keep supernatant (wash). Be careful not to remove the resin.
5. Elute recombinant protein twice with 300 μL buffer with 300 mM imidazole (incubate 5 min each time before spinning 3 min, $1,200 \times g$ at 4°C). Elution sample is obtained. Keep sample for PAGE-SDS or Western Blot.
6. Keep elution pools at 4°C for further use.

3.4 Analysis of First Round: Different Types of Additives. Alternative and Less Comprehensive Screen

The emphasis in this first screening is on checking additives that act by different mechanisms to suppress or avoid aggregation. For certain projects, this first run could be enough to determine the best conditions. An alternative screening use partially purified protein after the first capture step (*see* Note 20). This alternative screening, although faster and simpler, is less comprehensive. The best results

of these screens can be later applied to all the steps during medium- and large-scale purifications.

1. For each condition, run samples on SDS-PAGE: pellet, supernatant, unbound to IMAC, and eluted proteins (*see* Subheading 3.3). Analyze them by Coomassie staining.
2. Profile for the best additive: less target protein in the pellet and in the unbound fraction and higher protein concentration in the elution.
3. Keep best elution samples overnight (ON) at 4 °C.
4. Visual selection of non-turbid samples. Spin best samples 15 min 18,000×g at 4 °C and discard pellet. Only the best elution conditions are analyzed immediately by analytical or mini-analytical SEC, searching for the best monomeric/soluble aggregate ratio along time.
5. A simple but less informative option is to run SDS-PAGE after ON incubation at 4 °C, and spin: higher soluble protein along time, without any indication about the oligomeric conformation.
6. Check Western Blot only in case of low protein concentration or to verify the presence of the target and absence of cleavage products.
7. Final evaluation for this round: estimation of the influence of different additives groups on lowering insoluble aggregated proteins, allowing better binding to the IMAC resin, with the healthiest oligomeric conformation along time (Fig. 2).

3.5 Second Round of Optimization: Similar Additives or Combination of Additives from the First Round (Fig. 2)

In this optimization round, the emphasis is on finding alternative additives from the same group of the best additives from the first round and testing possible synergism of different additives with different modes of action.

1. Repeat small-scale IMAC purification using other additives of the same group as the best results from the first round (similar to [13]; for more information, *see* [12, 17]).
2. If trehalose gives the best results, try other osmolytes: 1 M TMAO, sorbitol or sucrose, 0.05 % polyethylene glycol 3,350 or 6,000.
3. If a detergent such as Tween 20 gives the best results, try 0.5–1 % of other surfactants like Nonidet P40, Tween 80, or Brij 35, or detergents used for crystallization of membrane proteins, octyl glucoside (*n*-octyl- β -D-glucoside) (OG) or *n*-dodecyl- β -D-maltoside (DDM).
4. If 0.5 % Zwittergent 3–14 gives the best results, try 1 M non-detergentsulfobetaines (NDSBs), 0.5 % 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), or Lauryl-dimethylamine *N*-oxide (LDAO).

5. If L-ArgHCl in the elution buffer is the best, try other amino acids as proline or a combination of 50 mM L-Arg with 50 mM L-Glu [20, 21].
6. Mix additives such as osmolytes and surfactants if both work or try other possible synergistic combinations.

3.6 Third Round: Lower Concentration of Additives (Fig. 2)

This optimization step is used for projects in which maximal decrease in additive concentration is important (e.g., detergents that can affect downstream applications, expensive additives, or undesirable chemicals such as urea or GdnHCl). Repeat low-scale IMAC purification using sequential dilutions of the target additive(s) (Fig. 3c). Alternatively, the additive concentration can be drastically reduced during elution or during later purification steps [7] (unpublished data).

3.7 Buffer Optimization Designed for Subsequent Purification Steps

This optimization step is performed when additives used for the first capture step are incompatible or undesirable in the following purification steps. Some other parameters not checked in the first purification step can be checked here: different pH, different salt concentrations (very important for ion or hydrophobic exchange columns), and other types of reducing agents or surfactants.

1. Dilute protein samples after first IMAC column 1:4 with a matrix of different buffers (*see* Subheading 2.5). The two main variables should be pH and conductivity. Additional additives and reducing agents can be added according to previous results.
2. Keep ON at 4 °C (alternative: experimental stresses; *see* Note 21).
3. Spin 20 min at 18,000 × *g* and 4 °C. Run SDS-PAGE or perform Western Blot on supernatants.
4. Profile of best conditions: most protein in the supernatant after long incubation at 4 °C.
5. Only the best conditions are analyzed immediately by analytical or mini-analytical SEC, searching for the best monomer/soluble aggregate ratio.

3.8 Set Up Concentration Limit and Best Stability/ Storage Conditions

For many biochemical and structural studies, there is a need for highly concentrated protein. Reaching such concentrations is a difficult task for prone-to-aggregate proteins. This screening is applied to purified protein in order to find the best buffer conditions for maintaining maximal protein concentration and long-term stability during storage. In this round, like in the previous round, different parameters should be tested including pH, salt concentration, and other types of reducing agents or surfactants.

1. Select concentrator (*see* Subheading 2.6). As a general rule, the pore size of the concentrator membrane should be two times smaller than the molecular weight of the protein. Select the concentrator volume size according to your needs.

2. Add some buffer to the concentrator and rinse the membrane. Use the concentrator immediately after washing and avoid drying the membrane. Always start with a small sample to determine the upper limit before concentrating the total amount of protein.
3. Spin according to the manufactures' instructions for a few minutes and check the protein concentration. If losses are higher than 20–30 %, check for protein concentration in the flow through. If protein is detected in the flow through, it may be that the unit is damaged or a smaller MWCO should be used.
4. Continue protein concentration by incremental steps. Take samples after each step. Aliquot the sample and keep part of the samples at $-80\text{ }^{\circ}\text{C}$ (*see Note 19*) and the rest of the sample ON at $4\text{ }^{\circ}\text{C}$.
5. Spin aliquots that were at $4\text{ }^{\circ}\text{C}$ (20 min $18,000\times g$, $4\text{ }^{\circ}\text{C}$). Run SDS-PAGE or check the protein concentration.
6. Profile evaluation of best conditions: highest protein concentration in the supernatant after long incubation at $4\text{ }^{\circ}\text{C}$.
7. Only the best conditions are then analyzed by analytical or mini-analytical SEC, looking for the best monomeric/soluble aggregate ratio.
8. Repeat same evaluation with aliquots keeps at $-80\text{ }^{\circ}\text{C}$ (*see Note 19*).
9. Use this information to concentrate and store your protein during scale-up.

4 Notes

1. Many laboratories use a simple protocol based on a small screening by SDS-PAGE to check the presence of soluble and insoluble proteins after cell lysis and centrifugation. We emphasize the importance of minimal presence of soluble aggregates as well as insoluble aggregates. The quality of the overexpressed product must be evaluated in order to minimize undesired aggregates during purification down the line. To reach this goal, we coupled small-scale expression and analysis by SDS-PAGE and analytical gel filtration for the optimal ratio of monomer/soluble aggregate in the bacterial lysates (similar to [22]). During expression, conditions must be found that give minimal presence of aggregates (both soluble and insoluble) and maximal yield of the native overexpressed protein.
2. Heat increases the kinetic energy of the protein chain, and this increase can break relatively weak H-bonds, electrostatic

interactions, and hydrophobic interactions, speeding up the aggregation process. pH change can affect the charge of acidic or basic functional groups in the protein and thus disrupt or create electrostatic repulsion that will alter the protein structure. Ionic strength can affect protein aggregation in different ways by reducing desired electrostatic interactions at high salt or increasing undesired electrostatic interactions at low salt. This can result in either stabilization or destabilization of proteins, or even denaturation [15]. This effect can differ for chaotropic or kosmotropic ions (mainly anions). Kosmotropic salts such as ammonium or magnesium sulfate stabilize the native protein state favoring protein–water interactions (so-called water-structure makers) [12, 13]. They are usually small ions with low polarizability and a bigger “salting-out” effect according to the Hofmeister series. Chaotropic salts, like magnesium chloride (with higher “salting-in” effect according to the Hofmeister series), are water-structure breakers and protein destabilizers. They can also inhibit protein–protein interactions by shielding charges and preventing stabilization by salt bridges [12, 13].

3. Factors that enhance protein stability interact mainly with the solvent. On the other hand, factors that suppress protein aggregation operate mainly by binding to the protein surface or by competitive binding to the interface that has the potential to destabilize the protein structure or cause aggregation [8].
4. Through the interaction of water molecules with osmolytes, water molecules are excluded from protein surface, thus stabilizing the native state of the protein with the smallest surface area [12, 23]. The addition of such cosolutes not only stabilizes many proteins but also deters ice formation, thus inhibiting the harmful effects of freezing on protein structure [12].
5. The amphiphilic polymer polyethylene glycol (PEG) is intensively used for protein refolding [24] and for protein stabilization by chemical modification (i.e., PEGylation) [25]. PEG interacts with the hydrophobic side chains that become exposed upon unfolding. Because of their high water solubility, low toxicity, and low antigenicity, PEGs are used in protein engineering to enhance refolding, assist in crystallization, increase water solubility, and prolong the blood circulation time of proteins [26]. Polyvinylpyrrolidone (PVP), a similar amphiphilic polymer, is applied in pharmaceutical products due to its low toxicity [15].
6. H-bonding agents, such as urea or GdnHCl, interfere with intramolecular interactions mediated by non-covalent forces such as hydrogen bonds, van der Waals forces, and hydrophobic effects. High concentration of these additives can lead to

protein unfolding by either a direct interaction with the protein [15] or an indirect effect on the surrounding water structure. Most likely, these two mechanisms are not mutually exclusive [21]. At low concentration (0.5–2 M), they act as chaotropic agents.

7. The mode of interaction between L-ArgHCl and proteins is still under extensive investigation [8, 14]. L-ArgHCl as an aggregation suppressor during refolding was first reported in a patent application [27]. Stepwise decrease of denaturant concentration in combination with the addition of L-ArgHCl is a conventional method for protein refolding [28]. It is also a versatile additive for protein formulation and affinity column chromatography [15]. It was shown to reduce nonspecific protein binding in SEC, to facilitate elution of antibodies from protein A columns, to enhance elution of resin-bound proteins, and as a solvent for elution in hydrophobic interaction chromatography (HIC) and ion exchange chromatography (IEC) [14].

A well-known synergistic enhancement of protein solubility is achieved by the combination of L-ArgHCl and L-glutamic acid (L-Glu). They interact with oppositely charged residues on the protein surface and mask the surrounding exposed hydrophobic patches [20, 21]. Only 50 mM of each compound are necessary, instead of the high concentrations (around 0.5–1 M) of L-ArgHCl alone. The mixture can be added to eluted protein after the first IMAC column and to all subsequent buffers.

8. Bi-tyrosine formation as a consequence of tyrosine oxidation is a chemical modification that can stimulate aggregation [5]. Oxygen scavengers such as methionine or sodium thiosulfate can avoid this aggregation [6].
9. Reducing agents must be used during extraction and purification if cysteines in the target protein are predicted or known to be free. This would prevent protein aggregation by inhibiting the formation of nonnative disulfide bonds. The most common reducing agents are dithiothreitol (DTT), β -mercaptoethanol (BME), or tris-(2-carboxyethyl) phosphine hydrochloride (TCEP). TCEP is a non-thiol and odorless compound, stable in aqueous solutions, and resistant to air oxidation. Unlike DTT, TCEP retains its reducing ability at acidic pH and at pH above 7.5 [29].

It is best to use BME during IMAC purification, since DTT or TCEP are incompatible with many of the IMAC resins. Using 5–15 mM, BME can avoid the formation of nonnative disulfide bonds. In other chromatographic procedures, BME can be replaced by other reducing agents.

No reducing agents must be used if only disulfide bonds are predicted. A problematic crossroad is a mixture of free cysteines and disulfide bonds in the same protein target. Our approach in this case is not to use reducing agents at all, or a very low BME concentration (2 mM), as a compromise solution.

There are several websites that can predict the bonding state of cysteines on proteins, such as Cyspred (http://gpcr.biocomp.unibo.it/cgi/predictors/cyspred/pred_cyspredcgi.cgi), DiANNA (<http://clavius.bc.edu/~clotelab/DiANNA/>), and DISULFIND (<http://disulfind.dsi.unifi.it/>).

10. Polysorbate 80 (polyoxyethylene sorbitan monooleate) and polysorbate 20 (polyoxyethylene sorbitan monolaurate) are surfactants that are widely incorporated in marketed protein pharmaceuticals. Used in the 0.0003–0.3 % range [8], they are reported to suppress aggregation upon agitation, shaking, freeze-drying, and freeze-thawing processes and can prevent protein adsorption at solid surfaces [6].

Poloxamers like the triblock copolymers of polyethylene oxide–polypropylene oxide–polyethylene oxide (PEO–PPO–PEO) or commercially available poloxamers such as Pluronics® or Synperonics™ are used in pharmaceutical formulations [30]. Poloxamer 188 (BASF Pluronic® F68) is widely used for the large-scale production of mammalian cell culture, especially when bioreactors are used to amplify a cell population [6].

Non-detergent sulfobetaines (NDSBs) are very good aggregation suppressors. They have a short hydrophobic group and a hydrophilic sulfobetaine head group, which is a zwitterion over a wide pH range. NDSBs do not behave like detergents, since their hydrophobic group is too short to form micelles even at concentrations as high as 1 M. This property allows them to be easily removed by dialysis. Moreover, they weakly bind proteins. All these reasons make them sometimes more useful than detergents [15].

11. Buffer conditions can potentially alter protein conformation or activity. These effects can vary at different cosolvent concentrations, using different cosolvents from the same family, changing protein concentration, or depending on the protein purification stage.

Screening of 144 additive conditions for increasing the solubility of recombinant proteins expressed in *E. coli* was recently described [7]. The classical additives gave the best results: trehalose, glycine betaine, mannitol, L-ArgHCl, potassium citrate, CuCl₂, proline, xylitol, NDSB 201, cetyltrimethylammonium bromide (CTAB), and K₂PO₄.

12. An easy alternative aggregation test is the visual observation of turbidity as a result of precipitation. This can be performed by

observing aggregates under a microscope (*Giladi, O., 2012 Rational optimization of protein stability. P4EU Workshop on Protein Purification*) or by optical observation at different wavelengths (340, 490, or 600 nm). These approaches, although fast and easy to perform, require large volumes of concentrated protein (to allow screening by buffer dilution) and can only detect highly insoluble and very large protein aggregates, while soluble aggregates remain undetected.

There are more laborious analytical methods to check for soluble aggregates. The most popular of these methods is SEC [22]. Other methods like CD or light scattering (LS) are not always available in all laboratories, and their results are more difficult to interpret. Analytical ultracentrifugation is the most accurate, but it is very expensive. Native gels are much cheaper, but need to be optimized for each protein and do not give an analytical result.

A filter-based aggregation assay used on crude cell extract or partially purified proteins was described [12]. After incubation in different buffers and under different conditions, the soluble, non-aggregated protein was separated from the big aggregates. This was performed using little ultracentrifugation devices where the MWCO was selected such that soluble protein was allowed to pass through the filter, while aggregate forms were retained. Analysis was done by SDS-PAGE or Western blotting [12].

Another method, used mainly for protein characterization for crystallography or NMR, is the fluorescence-based thermal shift (ThermoFluor) assay. An environmentally sensitive dye, Sypro Orange, is used to monitor the thermal stability of a protein. This assay can be used to investigate the effect of factors (buffers, additives, or ligands) on protein stability [31, 32]. RT-PCR machines with fluorescent detectors are used to compare shifts of T_m (midpoint of the unfolding transition on the melting curve).

There are several commercial assays with similar approach. All these can be employed to streamline protein processing and optimize formulation procedures:

- OptiSol™ Protein Solubility Screening Kit (Dilyx Biotechnologies) based on a filtration assay.
- ProteoStat™ protein aggregation assay (Enzo-Life Sciences or BioTek), using Thioflavin T as a fluorescence dye and a multi-mode microplate reader.
- Optim1000 (Avacta) combines fluorescence and static light scattering technologies.

As is described in the methods section, we prefer to use standard SDS-PAGE to select the best buffer and additives (electrophoresis of supernatant after ON incubation with

different additives), followed by SEC or mini-SEC for a more precise analysis. SEC completes the information about protein purity with information about the oligomerization state and allows a rapid estimation of the presence and amount of soluble aggregates, although larger aggregates seem to be lost in the pre-column filters [22]. Since aggregation is time dependent, we use ON incubation as a relative compromise. In addition, SEC can be coupled in-line to a light scattering device to measure the absolute molar mass, size, and shape of macromolecules in solution. Although not using a high amount of protein, the main disadvantage of SEC is that it is time consuming for nonautomated laboratories.

13. In a recent report, the first buffer selection was performed by the type of chemical that best improves solubility, followed by identifying the optimal chemical and its most effective concentration [13]. The report describes a filter-based aggregation assay used on crude cell extract to rapidly identify buffers that maintain protein solubility for purification and subsequent assays (*see Note 12*). A similar work was published some years ago with a very good table of agents that may promote protein solubility [12]. In spite of its simplicity, this approach yields less information regarding optimal purification conditions.
14. The isolation and purification of a tagged protein can be achieved by using a cheap and convenient affinity column that can yield tagged protein with 70–90 % purity following a single-capture step. Further purification is done by ion exchange, hydrophobic exchange, size exclusion chromatography, and the new mixed-mode chromatography columns (*see Note 16*) in order to achieve a higher degree of purification, which is often required for downstream applications.

Ion exchange chromatography is essential as an intermediate step for separating target proteins from protein contaminants such as chaperones and other host cell proteins. It also allows separating the target protein from heterogeneously folded forms that are a consequence of the expression and purification conditions used and from heterogeneous post-translational modifications. Sometimes ion exchange chromatography does not sufficiently separate the impurities, and additional chromatographic methods are required. These should be based on different principles, such as hydrophobic exchange, mixed mode, or hydroxyapatite. SEC is often recommended as a final purification step in order to eliminate protein contaminants and low molecular weight molecules and to obtain a homogeneous oligomeric form [33].

15. For some projects, we found that changing the order of the purification steps gave better results. This way the soluble aggregates were eliminated after the first capture step by SEC

before tag cleavage by specific proteases. SEC increased the purity of the protein and adjusted the initial buffer conditions for next columns. Long cleavage incubation times can be circumvented by increasing the protease concentration.

16. SEC is the method of choice to separate different oligomers. Symmetric elution profiles are characteristic of homogeneous proteins, whereas asymmetric profiles reflect nonhomogeneous, partially aggregated samples or large aggregates if eluting in the void volume of the chromatogram (or when the column is in poor condition) [34]. Recently, a great effort has been done to produce resins with high capacity and high flow rates, to be used for separating recombinant proteins from aggregates. Since these operate on a “mixed-mode” mechanism, based on a combination of electrostatic and hydrophobic properties of the proteins and ligands, they are called “multimodal” or “mixed-mode” resins. Examples of them are Capto adhere or Capto MMC (GE Healthcare); HEA, PPA, and MEP HyperCel (PALL); MX-Trp-650 M (Tosoh); Eshmuno HCX (Merck); and Hydroxyapatite (BioRad).

We have observed several times that high selective ion exchange columns can also separate different oligomeric states (unpublished data).

17. For prone-to-aggregate proteins, the ratio of lysis buffer to cell mass is extremely important and can lead to aggregation before the first purification step. We suggest to use at least twice or more lysis buffer for the same cell mass (1:5 to 1:10 of initial culture).

For some difficult projects, we preferred to use a batch binding of the crude lysate to the resin, in order to avoid the aggregation of the protein in the upper part of the column during loading. An alternative option is to use an excess of resin to avoid molecular crowding, although this approach can compromise the purity of the final product. A similar approach is used for purification of membrane proteins.

18. Since proteins are concentrated in the upper side of the columns during all chromatographic procedures except SEC, it happens that proteins with an extreme tendency toward aggregation start to precipitate immediately after elution. A small volume of buffer can be added to the collection tubes in order to obtain an immediate dilution of the protein and avoid or inhibit aggregation.
19. It is prudent to use a small sample to examine the stability of the protein for both protein concentration and freeze-thaw cycles before processing the entire batch. Be aware that during ultrafiltration (centrifuge-driven filter devices with adequate MWCO) a local over-concentration and irreversible precipitation

or aggregation of the protein on the filtration membrane can take place [34]. Small aliquots should be frozen in liquid nitrogen and then stored at -80°C to avoid damaging freeze-thaw cycles. Moreover, aliquots should always be thawed on ice [34].

20. Alternative screening evaluation: low-scale IMAC purification is performed without *additives*. The eluted protein is diluted 1:4 in buffers with different *additives* (concentrate protein with disposable ultrafiltration devices if the eluate is not concentrated enough). Then proceed to **step 3**, Subheading **3.4**.
21. Experimental stresses: *OptiSol™ Protein Solubility Screening Kit Application Manual* http://www.dilyx.com/protein_solubility_screen_home2.

Acknowledgment

AF is supported by a starting grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC Grant agreement n° 203413 and by the Minerva Center for Bio-Hybrid complex systems.

References

1. Chi EY, Krishnan S, Randolph TW et al (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* 20: 1325–1336
2. Berkowitz SA, Engen JR, Mazzeo JR et al (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat Rev Drug Discov* 11:527–540
3. Philo JS, Arakawa T (2009) Mechanisms of protein aggregation. *Curr Pharm Biotechnol* 10:348–351
4. Narhi LO, Schmit J, Bechtold-Peters K et al (2012) Classification of protein aggregates. *J Pharm Sci* 101:493–498
5. Cromwell ME, Hilario E, Jacobson F (2006) Protein aggregation and bioprocessing. *AAPS J* 8:E572–E579
6. Lee HJ, Mcauley A, Schilke KF et al (2011) Molecular origins of surfactant-mediated stabilization of protein drugs. *Adv Drug Deliv Rev* 63:1160–1171
7. Leibly DJ, Nguyen TN, Kao LT et al (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One* 7:e52482
8. Ohtake S, Kita Y, Arakawa T (2011) Interactions of formulation excipients with proteins in solution and in the dried state. *Adv Drug Deliv Rev* 63:1053–1073
9. Bandyopadhyay A, Saxena K, Kasturia N et al (2012) Chemical chaperones assist intracellular folding to buffer mutational variations. *Nat Chem Biol* 8:238–245
10. Singh LR, Dar TA, Rahman S et al (2009) Glycine betaine may have opposite effects on protein stability at high and low pH values. *Biochim Biophys Acta* 1794:929–935
11. Street TO, Bolen DW, Rose GD (2006) A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci U S A* 103:13997–14002
12. Bondos SE, Bicknell A (2003) Detection and prevention of protein aggregation before, during, and after purification. *Anal Biochem* 316:223–231
13. Churion KA, Bondos SE (2012) Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification. *Methods Mol Biol* 896:415–427
14. Arakawa T, Ejima D, Tsumoto K et al (2007) Suppression of protein interactions by arginine: a

- proposed mechanism of the arginine effects. *Biophys Chem* 127:1–8
15. Hamada H, Arakawa T, Shiraki K (2009) Effect of additives on protein aggregation. *Curr Pharm Biotechnol* 10:400–407
 16. Kissmann J, Joshi SB, Haynes JR et al (2011) H1n1 influenza virus-like particles: physical degradation pathways and identification of stabilizers. *J Pharm Sci* 100:634–645
 17. Lebediker M. Protein purification facility. http://wolfson.huji.ac.il/purification/Protocols/Additives_Folding.html
 18. Arnau J, Lauritzen C, Petersen GE et al (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif* 48:1–13
 19. Nallamsetty S, Waugh DS (2007) A generic protocol for the expression and purification of recombinant proteins in *escherichia coli* using a combinatorial his6-maltose binding protein fusion tag. *Nat Protoc* 2:383–391
 20. Golovanov AP, Hautbergue GM, Wilson SA et al (2004) A simple method for improving protein solubility and long-term stability. *J Am Chem Soc* 126:8933–8939
 21. Shukla D, Schneider CP, Trout BL (2011) Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. *Adv Drug Deliv Rev* 63:1074–1085
 22. Sala E, De Marco A (2010) Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography. *Protein Expr Purif* 74:231–235
 23. Timasheff SN (1998) Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. *Adv Protein Chem* 51:355–432
 24. Lee LL, Lee JC (1987) Thermal stability of proteins in the presence of poly(ethylene glycols). *Biochemistry* 26:7813–7819
 25. Roberts MJ, Bentley MD, Harris JM (2002) Chemistry for peptide and protein pegylation. *Adv Drug Deliv Rev* 54:459–476
 26. Muraoka T, Adachi K, Ui M et al (2013) A structured monodisperse peg for the effective suppression of protein aggregation. *Angew Chem Int Engl* 52:2430–2434
 27. Rudolph R, Fisher S (1990) Process for obtaining renatured proteins. US Patent 4,933,434
 28. Umetsu M, Tsumoto K, Hara M et al (2003) How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system. Spectroscopic evidence for highly efficient refolding of a single-chain fv fragment. *J Biol Chem* 278:8979–8987
 29. Willis MS, Hogan JK, Prabhakar P et al (2005) Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions. *Protein Sci* 14:1818–1826
 30. Kabanov AV, Lemieux P, Vinogradov S et al (2002) Pluronic block copolymers: novel functional molecules for gene therapy. *Adv Drug Deliv Rev* 54:223–233
 31. Nettleship JE, Brown J, Groves MR et al (2008) Methods for protein characterization by mass spectrometry, thermal shift (thermo-fluor) assay, and multiangle or static light scattering. *Methods Mol Biol* 426:299–318
 32. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2:2212–2221
 33. Lebediker M, Danieli T (2011) Purification of proteins fused to maltose-binding protein. *Methods Mol Biol* 681:281–293
 34. Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium et al (2008) Protein production and purification. *Nat Methods* 5:135–146