FEBS Letters 588 (2014) 236–246

journal homepage: www.FEBSLetters.org

Review

Production of prone-to-aggregate proteins

Mario Lebendiker *, Tsafi Danieli

Protein Expression and Purification Facilities, Wolfson Centre for Applied Structural Biology, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem, Israel

ARTICLE INFO

Article history:
Received 7 October 2013
Revised 30 October 2013
Accepted 31 October 2013
Available online 6 November 2013

Edited by Wilhelm Just

Keywords:
Intrinsically disordered protein
Intrinsically disordered region
Induction condition
Fusion protein
Chaperone
Protein aggregation
Protein storage
Protein concentration
Stabilizer
Aggregation suppressor
Chaotrope
Kosmotrope
Buffer condition
Aggregation analysis

ABSTRACT

Expression of recombinant proteins in *Escherichia coli* (*E. coli*) remains the most popular and cost-effective method for producing proteins in basic research and for pharmaceutical applications. Despite accumulating experience and methodologies developed over the years, production of recombinant proteins prone to aggregate in *E. coli*-based systems poses a major challenge in most research applications. The challenge of manufacturing these proteins for pharmaceutical applications is even greater. This review will discuss effective methods to reduce and even prevent the formation of aggregates in the course of recombinant protein production. We will focus on important steps along the production path, which include cloning, expression, purification, concentration, and storage.

1. Introduction

Aggregation is a complex process that originates by several different mechanisms [1–3]. Aggregates can be formed from self-association of the native conformation, or by structurally altered states. Aggregation can be typically induced by nucleation of a few proteins, which form small and soluble aggregates; these then serve as nucleation foci for the subsequent growth of larger insoluble aggregates. The nucleation-growth process can increase with time, temperature, protein concentration, and other parameters. Importantly, an extended lag phase can abruptly precede the formation of the large insoluble aggregates [1].

The intrinsic properties of proteins could be responsible for mediating the aggregation step when expressed in *E. coli*. One major example is the family of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) that belong to a large family of proteins possessing amino acid regions that lack a stable tertiary structure [4,5]. Their dynamic and flexible conformations readily tend to aggregate.

However, the intrinsic properties of proteins are not the sole factors contributing to the tendency of heterologous proteins to form aggregates. Factors related to the expression and to the purification conditions can play an important role in the misfolding of proteins. At the expression level, incompatibility of the bacterial machineries to fold proteins of eukaryotic origin, which include coupled transcription-translation mechanisms, the lack of suitable chaperones and post-translational modifications, as well as the absence of compartmentalization, may also contribute to the aggregation process. At the purification level, the physicochemical conditions surrounding of the protein, concentration and many others factors greatly influence folding. Numerous strategies have been developed to minimize protein aggregation and enhance their solubility. These include the following:(i) Developing procedures to tightly control the expression of the proteins using specialized promoters; (ii) Attaching the protein to solubility-enhancing fusion proteins; (iii) Developing specialized bacterial host strains; (iv) Screening for specific growth and induction conditions; (v) Considering practical methods to reduce aggregation during the purification steps; and (vi) Screening for suitable buffer conditions for protein purification.

* Corresponding author. Tel.: 972 2 658 6920; fax: 972 2 658 5735.
E-mail address: mario.l@mail.huji.ac.il (M. Lebendiker).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.10.044

Partial solutions can be found by optimizing each of these methods singly. However, we believe that developing a strategy that combines and integrates the optimization of all these methods simultaneously will maximize the potential of soluble protein production.

2. Factors that influence aggregation during the expression stage

2.1. Expression vectors

2.1.1. Choice of promoters

The first steps for successful expression of soluble protein start with cloning of the target gene into an expression vector containing a tightly regulated promoter. Tight regulation of transcription allows the expression to be carried out in a controlled environment, enabling not only the production of the target protein under optimal conditions, but also improvement of reproducibility and easier scale-up of the production conditions. This regulation can be further improved when the expression vector contains an origin of replication with low-copy number. One of the most popular systems used in numerous research and pharmaceutical applications are the T7-promoter-based vectors (commercially available from companies such as BD Novagen, NEB, and Invitrogen). These expression vectors contain a T7 promoter that is not recognized by the cellular RNA polymerase; therefore, they prevent leaky expression in strains that do not contain an exogenous T7 RNA polymerase gene. Specialized expression strains were developed for these systems, containing a chromosomal copy of the T7 RNA polymerase gene (DE3) under various promoters [6]. Inducing the expression of the T7 RNA polymerase results in the subsequent induction of a target gene cloned under the T7 promoter. This system has a large variety of vectors and host strains for various uses that will be discussed in detail in the following sections. Other tightly regulated vectors contain promoters such as the araC promoter, induced by arabinose [7], and the cspA promoter, induced by a temperature shift to 15 °C in the pCold vectors manufactured by TAKARA [8]. Cloning the target gene under such tightly regulated promoters allows screening under various environmental conditions and identification of the exact and reproducible conditions needed for enhancing protein solubility. Despite the obvious advantage of using powerful promoters that may lead to product accumulation of up to 30% of the total cellular protein, over expression can often drive the protein towards aggregation. In some cases there is an advantage in using promoter that can be fine-tuned, such as the araC, or even weak promoters such as the lac promoter, which allows a slower accumulation of correctly, folded proteins. There are cased in which best results can be obtained without induction, with only low levels of leaky expression accommodate optimized conditions for soluble protein production. However, this strategy is sometimes difficult to scale up due to variation in media formulations. In conclusion, the ability to control expression levels is a key element in the choice of bacterial expression vectors [9,10].

2.1.2. Choice of fusion proteins

Fusion proteins have the best success rate in improving the solubility of target proteins in *Escherichia coli* [11], and are often utilized to simplify the isolation using either their intrinsic properties, or an additional small affinity tag. However, it is difficult to accurately predict the effect of the various fusion partners on the solubility and expression profile of specific targets [11–13]. Therefore, it is often necessary to screen a battery of fusion constructs to determine which is most suitable. Despite the advantage of parallel screening of a large collection of tags, in recent years many protein production facilities have reduced this intensive screening to a selected few favorable fusion tags, thus avoiding massive high throughput screening (HTS) platforms, by matching the fusion protein's qualities to the requirements of the target protein partner. Among the more popular solubility-enhancing fusion proteins are the maltose-binding protein (MBP) [14]; thioredoxin (TrxA) [15]; nutilization substance A (NusA) [16]; small ubiquitin-related modifier (SUMO) [17,18], gluthatione S-transferase (GST), and several hyper-acidic short protein fusion tags [19].

MBP is the most studied solubility enhancer, and accumulating evidence suggests that it serves as a passive participant in the folding process; it acts as a stabilizer of partially folded target proteins, until spontaneous or chaperone-mediated folding occurs [20]. Despite MBP's considerable size (44 kDa), its high expression level, combined with its efficient purification options (dextrin Agarose columns or IMAC chromatography when a His-tag is added to the N-terminus), makes it a suitable candidate for solving a wide range of aggregation problems, and it should be included in most fusion-protein expression screens [21]. MBP is suitable for tagging relatively smaller proteins (up to around 40 kDa) because the bacterial machinery is less efficient when large proteins (over 90 kDa) are produced, and this tends to result in low productivity and partially truncated protein forms.

Another large fusion protein that facilitates solubility similarly to MBP is NusA, a 55 kDa protein, highly soluble in *E. coli* [22]. Although this protein enhances the solubility of target proteins, sometimes with a higher efficiency than MBP, its large size and tendency to adhere to the target protein after attempts to cleave off the tag constitute a considerable disadvantage. An additional noteworthy fusion partner is the SUMO protein. It has solubility enhancement effects similar to MBP, and is gaining popularity owing to its accumulating successful results, its small size, and an efficient and highly specific tag-removal procedure with SUMO protease. The TrxA protein, which is about 11 kDa in size, can be used with larger target proteins, but it is most suitable for enhancing the solubility of proteins that contain disulfide bonds [23]. GST has been one of the most traditional tags used for many years, for over-expression and enhanced solubility. In recent years it has mainly been used as a popular tag for pull-down assays and protein–protein interaction studies. However, despite its great advantage in these applications, according to our personal experience and that of other protein production laboratories participating in the Protein Production and Purification Partnership in Europe (P4EU), GST seldom contributes to the solubility of its fused target protein in *E. coli*, and can cause pre-mature termination of the polypeptide chain, or even enhance aggregation, owing to its dimeric form.

Other solubility enhancers, like acidic fusion partners, act as "electrostatic shields", reducing the probability of aggregation via electrostatic repulsion between highly charged soluble polypeptide, thus allowing adequate time for correct folding. In addition, these solubility enhancers might directly act as intramolecular chaperones by participating in native folding of the target proteins [19]. This family of fusion proteins includes the lipoyl domain from *B. Stearothermophilus* E2p [24], consisting of a short acidic 109 amino acid tag (pI: 4.53 and MW: 11 994.3 Da). The lipoyl domain fusion tag containing the N-terminus His tag and an optimized Tobacco Etch Virus (TEV) protease cleavage site (termed HLT-tag) is used at our facility as the preferable fusion protein for enhancing the solubility of IDR proteins [25–27]. This tag is best suited for NMR studies and is highly resistant to proteases (received from Dr. Mark Allen: MRC-CPE, Cambridge, England).

Other noteworthy fusion partners are the N-Domain of *E. coli* phosphoglycerate kinase, a 22 kDa protein domain, recently reported by the Lee group to enhance solubility of several prone-to-aggregate proteins [28], and two additional tags: the modified bacteriophage T7 protein kinase and the *E. coli* Skp chaperone,

analyzed by the Esposito group [29]. These tags were tested in combination with several prone-to-aggregate targets, and exhibited enhanced solubility of the target proteins even after cleavage of the fusion tags. There are additional fusion proteins that may support solubility in disulfide bond-containing proteins, such as DsbA [12,30], which is used for periplasmic secretion and cytosolic expression, and the DsbC tag, which facilitates disulfide bond formation [31]. These options and many others should be considered either in a HTS format or for research projects requiring specific protein characteristics.

Fusion proteins and tags often have to be removed from the target protein owing to their potential tendency to interfere with downstream applications. Hence, efficient removal of the tags plays an important role in their selection [32]. Proteases such as thrombin and factor Xa have traditionally been used for many years; however, in recent years these proteases have been replaced by more specific, stable, and easy-to-produce proteases that are active at low temperatures. These proteases include the Tobacco Etch Virus (TEV), Rhinovirus 3C Protease (i.e. PreScission™), SUMO protease (which specifically recognizes SUMO protein), and enterokinase. Comprehensive reviews on proteases can be found in [32–35].

Despite the obvious advantages of using fusion proteins, three main drawbacks must be emphasized: (1) pharmaceutical companies are often reluctant to use them, due to the high cost of the tag removal process, and the extended purification steps needed to avoid the risk of contaminating the protein preparations with a partially cleaved product. (2) Most of the proteases leave a few additional amino acids at the N-terminus of the fused protein after cleavage. This may interfere with protein activity, immunogenicity, and other properties. (3) The main drawback of cleaving of fusion proteins is that cleavage is often followed by loss of solubility of the target protein. Circumventing this last drawback will be discussed in section 3 of this review.

2.1.3. Gene cloning

In recent years, improvements in technology have enabled cost-effective production of synthetic genes, and consequently, ordering codon-optimized genes for heterologous expression in *E. coli* has become a standard practice. Codon-optimized genes often increase the production levels and may also prevent premature termination of the elongated polypeptide [36]. Nevertheless, rare codons play a crucial role in regulating protein production and in protein folding [37]. Hence, we postulate that the use of codon-optimized genes might sometimes result in proteins that are more likely to aggregate. Although there is still no direct evidence to support this speculation with regard to genes expressing in *E. coli*, it might be worthwhile to consider these variables when developing future codon optimization algorithms.

Taking all this into consideration, we believe that the benefits of using synthetic codon-optimized genes outweighs their drawbacks, and enable a simple and efficient cloning method for most genes, which simplifies the cloning of complex and custom-designed constructs.

Another strategy employed when attempting to express a large multi-domain protein is to design a series of truncations to produce smaller, single domains that are easier to express in a soluble form. This can be achieved by designing primers to amplify specific domain boundaries and by expressing multiple domains in cell-free arrays or in parallel, cloning them into expression vectors. This strategy can be employed mainly for HTS platforms, and it is most beneficial when combined with an enzymatic activity assay that can detect active domains in a HTS platform. In some cases, structural and molecular displacement data may allow small-scale screening of rationally designed domains. When no structural data are available, there are several on-line tools for predicting soluble domains. However, the precise domain boundaries can vary, depending on the prediction method used [38]. Domain selection has limited uses; it increases the screening cost, and can be employed only when smaller domains are suitable for the required applications.

2.2. Bacterial host strains

Parallel screening of expression in various bacterial strains can often result in enhanced solubility in specific strains. Sometimes for reasons that are not completely understood, screening of expression in BL21-based strains and K12-based strains may yield a completely different solubility profile of the same protein. However, as in the case of selecting suitable fusion partners, HTS platforms are not always an available option, and therefore carefully selecting a small number of host strains may be sufficient to design an efficient screening procedure.

As in the case of tightly regulated promoters, there are considerable advantages of using strains that support tight regulation of induction and that prevent leakiness. Good examples of such strains are those that contain a chromosomal copy of the T7 RNA polymerase gene cloned downstream from a regulated promoter. These include strains such as BL21AI (Invitrogen), in which the induction of the T7 RNA polymerase is controlled by arabinose, and where leakage of the target protein is suppressed until suitable conditions are achieved in the culture. Other available host strains were created by mutating the lac permease (*lacY*) gene to allow uniform entry of IPTG into all cells. This produces linear concentration-dependent induction and enables one to maintain low concentrations of an inducer throughout the induction process (such as Tuner™ from Novagen). These specialized strains, which allow better control over environmental and internal conditions, increase the chances of enhancing solubility by reducing the translation rate and allowing proper folding of the protein in the bacterial coupled transcription-translation mechanism. However, there are cases when these changes are still insufficient for large-scale production of proteins that are prone to aggregate.

Several bacterial strains were developed to address the specific characteristics of aggregating proteins. The C41(DE3) and C43(DE3) strains, with proliferated intracellular membrane, were originally isolated by the Walker group [39] and were found to support higher production of active membrane proteins. A more recently developed strain, Lemo21 (DE3), contains mutations in the lacUV5 promoter governing the expression of T7 RNA polymerase [40–42], which was also designed to enhance the expression of active membrane proteins. Although this review does not directly address the aggregation of membrane proteins, our experience shows that these commercial strains (available from Lucigen and NEB) can successfully be utilized to enhance the solubility of not only membrane proteins but also hydrophobic and IDR proteins.

One of the most prominent causes for the formation of protein aggregation and inclusion bodies is the reducing environment in the cytosol, which does not allow the formation of disulfide bonds. One option to facilitate the correct formation of disulfide bonds is to secrete the protein to the periplasmic space; however, this will often result in a dramatic decrease in productivity. Specialized strains with impaired activity of the thioredoxin reductase (trxB) and glutathion reductas (gor) genes allow diminishing of the reductive pathways, and can support the formation of disulfide bonds in the cytosol [43–45]. These strains are commercially available from Novagen (Origami™). However, even in these strains the ratio between properly modified protein and aggregated protein is still low, especially under over-expression conditions. In a recent study [46], researchers demonstrated that adding a chromosomal copy of the disulfide bond isomerase, DsbC, which lacks its native signal sequence, facilitates achieving higher efficiency of disulfide

bond formation in cytosol, allowing better yields of the recombinant protein. This strain is currently commercially available from NEB. It is important to emphasize that most therapeutic proteins contain disulfide bonds [44]; hence, there is a growing need to develop more robust bacterial hosts that allow over-expression of correctly folded target proteins.

2.3. Expression conditions

Expression conditions and regulated induction play a crucial role in preventing aggregates during the production stage. The expression conditions consist of the following: induction temperature, induction duration, inducer concentration, and media formulation. Changes in the induction temperature may influence the elongation rate of the polypeptide chain, consequently inducing the collapse of the elongating polypeptide chain into ordered structural domains. Reducing the induction temperature often has a positive effect on the solubility, attributed to the slower elongation of the polypeptide chain. Slowing down this step allows bacterial chaperones and chemical additives to protect the elongated chain and it supports correct folding [47,48]. Inducer concentration affects the mRNA production and influences the local concentration of elongated polypeptide chains. Reduced concentrations of the inducer may prevent aggregation of the growing chains while they are still on the polysome. In addition, osmolyte chemical chaperones such as glycerol, proline, and sorbitol, when added in small concentrations to the growth media, may further impede the process, enabling a better-suited micro environment for proper folding during the peptide elongation step[49,50] [51].

2.3.1. Auto-inducing media

Traditional expression screens include parallel use of various media formulations differing in peptone concentration, yeast extract, and salts. In recent years, development of media formulations for auto-induced expression has proved to have a dramatic influence on protein production in general and on soluble product in particular. Auto-inducing formulations are based on the well-studied mechanism by which glucose suppresses induction by lactose. Together with metabolic balancing of pH and additional factors, specialized media were developed that allow reliable suppression and auto-induction of proteins of high density in bacterial cultures. The autoinduced media formulations are mostly compatible with T7-based expression systems, but they can also be adapted for arabinose induction and used with an extended range of bacterial expression systems [52]. These formulations allow parallel screening of numerous factors in combination with various media formulations, based on a simple inoculation procedure, and allow this media to be compatible for both large-scale HTS screening and for more limited screening platforms.

2.3.2. Screening media components

Including different media formulations in expression screens is a simple and often beneficial manipulation, although it is difficult to predict which formulation (whether the higher or lower nutrient concentration) will improve the expression levels and solubility of a target protein. A promising aspect of media formulation is currently under research in our laboratory: exploring the potential of specific *Saccharomyces cereviciae* peptones and yeast extract combinations, such as the Springer® yeast extract and yeast peptones to enhance the production of soluble proteins. It is difficult to predict which formulation combination will result in enhanced solubility; however, screening these extracts in combinations in an auto-inducing media resulted in enhancement of soluble protein production compared to conventional media formulations (unpublished data).

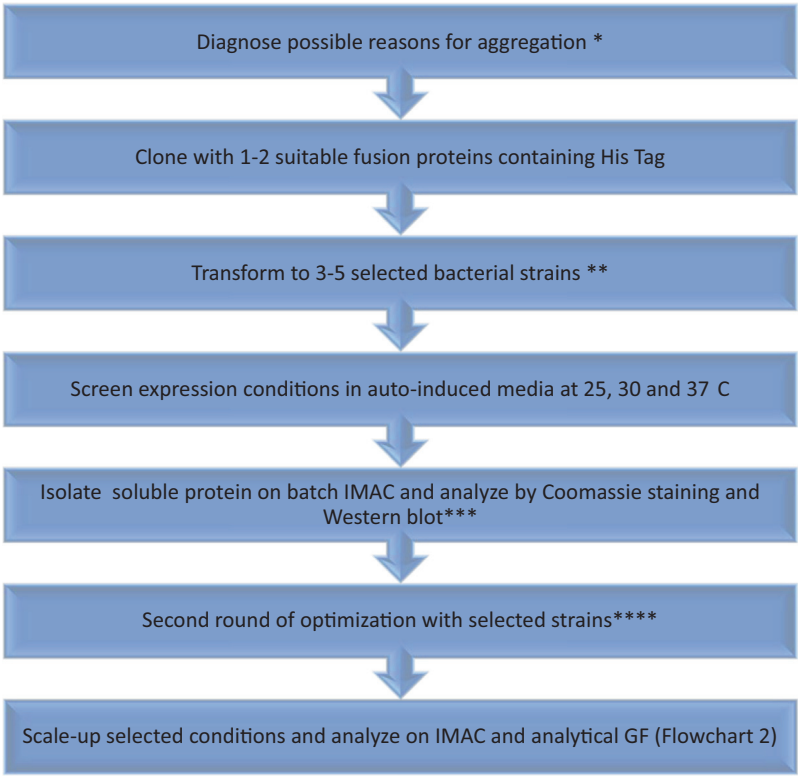
2.3.3. Induction and co-expression with chaperones

A general strategy to improve the native folding and solubility of recombinant proteins is to increase the cellular concentration of chaperones in the vicinity of the elongated polypeptide chain. There are two ways to achieve this: one is by treating bacterial growth with "heat-shock" prior to induction, to stimulate the production of heat-shock chaperones. This is usually done in our laboratory, combined with a low concentration of chemical chaperones: glycerol, potassium glutamate and benzyl alcohol, that elevates the cellular concentration of these viscous osmolytes, and enhances the effect [49,50]. Another more controlled methodology is to express with specific chaperones, for example: DsbA and DsbC, to promote disulfide bond formation in cytomlasmic or periplasmic space [53,54]. This strategy can also be combined with the addition of osmolytes. When lacking specific knowledge of suitable chaperones, screens for expression in the presence of combinations of chaperones can be conducted [55]. A comprehensive review examining the advantages and disadvantages of this approach can be found in a review by Kolaj et al. [56].

2.4. A combined approach for screening under solubility-enhancing expression conditions

As described in chapters 2.1 – 2.3, there are numerous vectors, bacterial strains, and environmental conditions that can be tested in order to determine the conditions for improving protein solubility. Many but not all parameters can be screened using HTS platforms. Some procedures, such as using unique expression systems and induction methodologies, should be carried out individually, in cases where the HTS format is not suitable or is too laborious. Most research laboratories focus on a single target, or targets from the same family of proteins. For this type of project we have developed a simple and cost-effective screen described in Flow chart 1. In this screen, the features of the individual target are first analyzed using databases and on-line bioinformatics tools. This first step will determine the choice of one or two fusion partners that will be attached to the N-terminus of the protein, in a T7 promoter-based vector. For example, if the protein contains IDRs, we will construct the gene with an N-terminus HLT tag and either a His-MBP or His-SUMO, depending on the size of the protein (i.e. proteins smaller than 40 kDa will be attached to MBP, whereas larger proteins will be attached to SUMO). We will then decide on several appropriate bacterial strains. Here, the standard BL21(DE3) and HMS174(DE3) strains should be added to the C41(DE3) and C43(DE3) strains. Parameters for selection of fusion partners and host strains are presented in Table 1. We usually transform two constructs into four host strains, inoculate single colonies into an auto-induced media array for 16 h at two different temperatures (usually 37 and 30 °C), lyse the samples, load on batch IMAC, and then analyze elutions by SDS–PAGE using Coomassie staining and Western blots. This procedure usually takes two days and enables one to select the best combination of fusion protein and expression strain. Additional parameters such as longer induction times, lower temperatures, and various media formulations can be added in the same platform. After selecting the most promising construct-host strain combination, we proceed by refining the screening to test the effect of inducer concentration, temperature, and induction duration. Batch IMAC elutions are analyzed by SDS–PAGE as in the first screening. The most efficient combination will be scaled up to medium scale growth (50–100 ml) and analyzed on IMAC and on analytical Gel Filtration (GF). This process covers the most influential expression aspects of protein solubility in less than a week from obtaining the cloned constructs (Flow chart 1).

This simple and modular screen follows the "20/80 rule", i.e., achieving 80% of the benefits with only 20% of the resources, while addressing the majority of solubility problems. It can easily be



Flow chart 1. Screening methodology for expressing prone-to-aggregate proteins *According to the literature, bioinformatics data analysis. **According to the requirements of the target protein, for example: expected disulfide bonds will include Shuffle and Origami strains in addition to standard strains such as BL21 (DE3) and HMS174. ***Coomassie staining allows semi-quantitative estimation of yield whereas Western blot analysis allows the detection of truncated and degraded byproducts. ****Screening using various inducer concentrations, temperatures, induction durations, and media formulation (heat shock: if required).

adjusted for many types of proteins, and when combined with the subsequent procedures for isolation and concentration, it proved to be extremely efficient for the vast majority of IDR, IDP, and other types of proteins prone to aggregate that were tested at our facility over the years [25–27,57–59].

3. Factors that influence protein aggregation during purification

Although protein solubility during expression is an essential prerequisite, aggregation problems can often arise at later stages during the purification process.

It is often tempting to try to solve solubility problems by decreasing protein concentration or by changing pH or salt concentrations. These simple solutions may dissociate the aggregates in cases where molecules are held together by relatively weak, non-covalent interactions. pH- or salt-dependent reversibility is indicative of equilibrium between the monomer and high-order forms [60]. However, such changes rarely affect most types of aggregation and we discuss here a more comprehensive approach that should be taken to deal with protein aggregation.

3.1. General issues

There are many protein stability issues that must be addressed starting at the very early purification steps until final storage. As can be seen in Table 2, there are general issues to consider: low purification times, low temperatures of purification, low protein concentrations at each step, prevention of mechanical or non-mechanical stresses (freezing, exposure to air, interactions with metal surfaces, etc.) and others.

3.1.1. Avoid high protein concentrations

A source of aggregation problems during processing could be the high protein concentration during different purification steps such as:

(1) The lysis step: this can be solved by using a higher ratio of lysis buffer to cell paste.
(2) Protein crowding at the top of the column during chromatographic loading: this can be solved by using a batch procedure or using an excess of resin.
(3) High protein concentration during elution: dilute proteins by adding buffer to collection tubes, or immediately after elution exchange the buffer by dialysis or buffer exchange columns.
(4) During concentrations: before any protein concentration optimize the buffer composition and procedure to be used before any protein concentration.

3.1.2. Work at low temperatures - Early elimination of protein aggregates

There is a possibility that the presence of soluble aggregates after the first IMAC column will trigger the insolubility of more protein molecules and decrease the total yield (unpublished data). Therefore, important points for prone-to-aggregate proteins are (a) Slow aggregation process by working at low temperatures. (b) Remove soluble aggregates as soon as possible by performing size exclusion chromatography (SEC) immediately after the IMAC purification and before tag cleavage (instead of the common strategy described by [22,33], where recombinant protein is first cleaved during dialysis, followed by negative IMAC and final polishing SEC).

Table 1
Common reasons for aggregation and the way to address them during the expression procedures.

A Cause of aggregations	B Vector	C Host Strain	D Growth conditions*	E Refs.
Lack of correct disulfide bond formation	Fuse with TRX, DsbA, and DsbC fusion partners. Clone with periplasmic secretion signals (pelB, OmpA)	Use Origami and Shuffle strains for cytosolic expression	Standard screen	[44–46]
Intrinsically disordered protein	Fuse with MBP, Sumo, and lipoyl-domain tag	Membrane-rich strains such as C41 and C43	Standard screen Heat shock with chemical chaperones	[25–27]
Hydrophobic protein	Fuse with solubility-enhancing proteins such as MBP, and SUMO	Membrane-rich strains such as C41 and C43	Standard screen Heat shock with chemical chaperones	[14,21]
No appropriate chaperones	Co-express with chaperone-containing vectors	Screen various BL21 and K12 strains	Standard screen Heat shock with chemical chaperones	[49–51,54]
Protein is natively directed to a sub-cellular localization	Remove the localization signal or replace the signal with a periplasmic secretion signal (pelB, OmpA)	Membrane-rich strains such as C41 and C43 or Lemo strain	Standard screen Reduce inducer concentration Induce at a high OD	[41]
Membrane protein	Generate and screen soluble domains	Use membrane-rich strains such as C41 and C43, or Lemo strain	Lower the induction temperature Reduce the inducer concentration Induce at a high OD	[41,42]
Protein is part of a complex	Fuse with large fusion proteins such as MBP. NusA co-expresses with a partner: a combination of 2–4 vectors for max 8 proteins	Screen tight induction-regulated strains	Heat shock with chemical chaperones	[49,50]

Column A specifies a list of reasons that can lead to protein misfolding during the expression steps. Columns B and C state changes in the parameters in the expression vector and the host cells, respectively. Column D describes various growth conditions, where "a standard screen" refers to the screening of variables such as temperature, the duration of induction, screening of various inducer concentrations, and screening of various media formulations. Each of the sections can be changed separately, or combined together with the others.

* Our standard heat-shock procedure involves the addition of 0.1% glycerol and 0.1 mM potassium glutamate. Detailed procedures for both standard screens and heat-shock procedures can be found in the following link: http://wolfson.huji.ac.il/expression/procedures/bacterial/Induction.Condition.Callib.new.htm.

3.1.3. Quick strategy of purification

Since the process time is one of the most critical points to consider, great emphasis must be given to designing a "quick strategy of purification", where pure protein must be produced and stored as fast as possible. To achieve this goal, intensive work must be carried out to optimize and fine-tune each purification step before scale-up. This will guarantee that the whole process can be achieved swiftly and smoothly.

3.2. Protein environmental issues

In addition to the general considerations, there are specific considerations involving environmental buffers: pH, ionic strength, cosolutes such as chaotropes and kosmotropes, osmolites and ligands, reducing agents, surfactants, and non-denaturative detergents (Table 2). Although many recommended additives are found in the literature, the type and optimal range for each protein is highly specific and the specific conditions must be fine-tuned for each project (Table 3). These environmental factors can either act by stabilizing proteins or by inhibiting aggregation and protein–protein interactions [61,62].

Owing to the difficulty of predicting the efficiency of each of these factors, the logical option is to test a big matrix of different buffers at diverse pHs and conductivities in the presence of different additives like those presented in Table 3. Leibly et al., using a screening methodology with 144 conditions, showed in a recent publication that only the classical ones give the best results: trehalose, glycine betaine, mannitol, L-arginine, potassium citrate, $CuCl_2$, proline, xylitol, NDSB 201, CTAB, and K_2PO_4 [63]. This confirms our experience that short lists of additives, tackling insolubility by different ways, and covering most of the aggregation mechanisms, considerably reduce costs and efforts. A similar short list and hierarchical approach of relevant buffer selection has been previously reported by the Bondos lab [61,64], where the first step is testing additives in each category, which is followed by a second screening of similar additives from the same family.

3.3. Methods for monitoring protein aggregates

Various experimental methods are routinely used to check aggregation. The initial step in most approaches is to test the protein sample using a combination of different buffers and additives (Table 3), and to incubate the reaction overnight at 4 °C (or under optional incubation conditions). These different screens can be used on purified or partially purified protein samples and analyzed using several methods.

Functional biological assays can be used when the target protein can be tested in a simple activity assay. However, this approach is not suitable for most proteins since such assays are not always available. However, most importantly, activity assays do not provide information regarding yield, oligomeric homogeneity, and protein purity.

Other simple methods such as visual observation of turbidity under a microscope [65] optical density at different wavelengths (340, 490, or 600 nm), turbidimetry, and analyzing spin-filtered samples on PAGE–SDS, although simple to perform, identify only big massive aggregates and cannot detect soluble aggregates. A filter-based aggregation assay using ultrafiltration devices was developed to identify soluble aggregates based on the MW cut-off of the monomer protein [64]. However, this approach is not very reliable and many artifacts can affect the final results.

Instruments such as circular dicroism (CD), light scattering (LS), and very accurate analytical ultracentrifugation (with low sample throughput) can identify the presence of soluble aggregates; however, these instruments are not always available in many laboratories, and the results are often not easily interpreted. Moreover, CD is incompatible with UV active substances that can be found in many buffers, while LS cannot always detect monomers in the presence of large quantities of aggregates (Sabine Suppmann, The Recombinant Protein Production Facility of the Max-Planck Institute in Munich, personal communication). Use of native gels can be a simple solution due to their low cost; however, these gels require laborious optimization for each protein and many artifacts can affect the final results.

Table 2
General and environmental considerations during protein purification steps.

Issues	
General issues	
Temperature	Heat increases the kinetic energy (energy of motion) of the protein chain. Excessive motion can break relatively weak H-bonds, as well as electrostatic and hydrophobic interactions. An increase in temperature accelerates chemical reactions such as oxidation and deamidation. Carrying out the purification process under low temperature conditions may reduce the tendency for aggregation
Time	Aggregation is a nucleation, time-dependent growth process. A "quick strategy of purification" must be designed to allow protein storage as soon as possible
Protein concentration	For prone-to-aggregate proteins, high protein concentrations in each of the purification steps can start or speed up the nucleation aggregation process: (1) Ratio lysis/cell (2) Protein concentration during each step (3) Overloading column (4) High protein concentrations during elution
Stresses	Freezing and thawing, exposure to air, interactions with metal surfaces, agitation, and additional mechanical or non-mechanical stresses might increase the tendency to aggregate. Minimizing these stresses may reduce aggregation
Protein strategy	Extended purification procedures, and the presence of soluble aggregates may accelerate protein insolubility. This nucleation process can be considerably avoided by early elimination of soluble aggregates (SEC or other chromatographic procedures like ion exchange or mixed mode chromatography), and a quick strategy of selection of a purification optimization procedure and arrangement of methods and columns
Protein environment	
Buffer type, pH, and salt concentration	*Salt concentration:* This can affect the aggregation of different proteins differently; reducing electrostatic interactions at high salt concentrations, or increasing electrostatic interactions at low salt concentrations. These changes can cause either the stabilization or destabilization of proteins, or even denaturation [71] *pH:* pH can greatly affect the protein structure. It can change the electric charge of acidic or basic functional groups on the protein as well as disrupt or create electrostatic interactions that will alter the protein structure. pH determines the electrostatic interactions through charge distribution on the protein surface [72]
Stabilizing (Kosmotropic) agents. Decrease aggregation by stabilizing the protein structure, avoiding partial unfolding. In some cases these agents can increase aggregation by enhancing protein–protein associations [61]	*Osmolytes (sugars, amino acids, polyols, etc.):* Decrease aggregation by stabilizing the structure of water-water interactions, which causes water molecules to favorably interact with macromolecules. This stabilizes intermolecular interactions [64,73–75 *Kosmotropic salts:* These salts have a higher salting-out effect according to the Hofmeister series. They act as a protein stabilizer (usually small ions, low polarizability), and as polar water-structure makers *Detergents:* Surfactants can stabilize proteins by two major mechanisms: (a) by settling at an interface, and preventing protein adsorption and associated surface activity loss and/or surface-induced aggregation or (b) by binding to hydrophobic patches of proteins and thus preventing a close approach and aggregation. Some surfactants may function according to only one of these mechanisms, whereas others may function according to both [3] *Ligands, inhibitors, etc.:* Can stabilize the protein structure
Aggregation destabilizing (Chaotropic) agents. May destabilize aggregates formed by proteins in or near the native state. But in other cases they can cause partial denaturation and consequently enhance aggregation [61]	*H bonding agents (urea and GuHCl):* For several proteins, low concentrations of these agents (up to 2 M) can destabilize aggregation by decreasing the net hydrophobic effect of hydrophobic regions by disordering water molecules adjacent to the protein [71]Higher concentrations of these agents can lead to unfolding of the protein chain by interfering with intra-molecular interactions mediated by non-covalent forces such as hydrogen bonds, van der Waals forces, and hydrophobic effects *Arginine:* The mode of interaction between Arginine (Arg) and protein [73] in is still under extensive investigation [73,76]. Arg as an aggregation suppressor during refolding was first reported in a patent application [77]; its stepwise reduction of denaturant concentrations in combination with the addition of Arg is the most conventional method for protein refolding [78]. Like GdnHCl, Arg interacts in a similar manner with the amino acid side chains and peptide backbones, suggesting that it has affinity for side chain groups, most significantly for aromatic side chains [76]. Arg is also a versatile additive for protein formulation and affinity column chromatography [71] *Reducing agents:* Mild reductants and mild oxidants can lead to changes in protein conformation that may alter the function of the protein. Mild reductants can break disulfide bonds and may lead to dissociation of parts of the protein chain(s) that are normally associated. Mild oxidants can cause the formation of disulfide bonds and may lead to association of parts of the protein chain that are normally not associated. A problematic crossroad is a mixture of free cysteines and disulfide bonds in the same protein. Our approach in this case is not to use reducing agents at all during this process, or at least at a very low BME concentration (2 mM) as a compromise solution *Chaotropic salts:* They have a higher "salting-in" effect according to the Hofmeister series. On one hand, they are water-structure breakers and protein destabilizers; but on other hand, they can reduce protein–protein interactions by shielding charges and by preventing the stabilization of salt bridges [61,64,79] *Detergents:* Some detergents disrupt hydrophobic interactions

Table 3
Additives Used to Stabilize Folding and to Prevent Aggregation Summary table of different publications [50,61–63,69,78–81] and from commercial websites (DILYX Biotechnologies OptiSol protein solubility screening kit application manual, HAMPTON: Solubility and stability screen).

Additive	Recommended initial concentration	Recommended concentration range
Sugars and osmolytes		
Glycerol	10%	0–40%
TMAO (trimethylamine N-oxide)	0.5 M	0–1 M
Glucose	0.5 M	0–2 M
Sucrose	0.5 M	0–1 M
Trehalose	0.5 M	0–1 M
Ethylene glycol	10%	0–60%
D-Sorbitol	0.5 M	0.2–1 M
Mannitol	2%	
Xylitol	0.5 M	0.2–1 M
Glycine betaine	1 M	
Amino acids and amino acid derivatives		
Glycine	250 mM	0.5–2 M
Arginine L-HCl	125 mM	0–2 M
Arginine ethylester	250 mM	0–500 mM
Proline	250 mM	0–1 M
Potassium glutamate	250 mM	0–500 mM
Arginine L-HCl + L-glutamic acid (L-Glu)	50 mM each	
Non-ionic detergents		
Nonidet P40 (NP40) or Triton X-100	0.01%	0–1%
Tween 80 or 20	0.1%	0–1%
DDM: *n*-dodecyl β-D-maltoside	0.1%	0.01–0.5%
Brij 56: polyoxyethylene cetyl ether	0.05%	
OG: Octyl glucoside (*n*-octyl-β-D-glucoside)	0.1%	0.01–0.5%
Zwiterionic detergents		
NDSB: non-detergent sulfo betaine	0.5 M	0–1 M
CHAPS: 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate	0.1%	0.01–0.5%
Zwittergent 3–14	0.1%	0.001–0.2%
LDAO: lauryldimethylamine *N*-oxide	0.1%	0.01–0.5%
Ionic detergents		
CTAB: cetyltrimethylammonium bromide	0.5%	
Sarkosyl: sodium lauroyl sarcosinate	0.05%	0.01–0.5%
SDS: sodium dodecyl sulfate		Up to 0.1%
Mild chaotrope agents and chaotrope salts		
Urea	0.5 M	0–2 M
Guanidine HCl	0.5 M	0–2 M
N-Methylurea	250 mM	Up to 2.5 M
N-Ethylurea	100 mM	Up to 2 M
N-Methylformamide	3–15%	
NaI	0.2 M	0–0.4 M
$CaCl_2$	10–50 mM	0–0.2 M
$MgCl_2$	10–50 mM	0–0.2 M
Mild and strong kosmotrope salts		
NaCl (weak)	300 mM	0–1 M
KCl (weak)	200 mM	0–1 M
$MgSO_4$ (strong)	100 mM	0–0.4 M
$(NH_4)2SO_4$ (strong)	50 mM	0–0.2 M
Na_2SO_4 (strong)	500 mM	0–0.2 M
Cs_2SO_4 (strong)	50 mM	0–0.2 M
Potassium citrate	100 mM	
Citric acid	50 mM	
Alcohols, polyols, polymers, polyamines, and others		
Ethanol	5–10%	Up to 25%
n-Penthanol		1–10 mM
n-Hexanol		0.1–10 mM
Cyclohexanol		0.01–10 mM
Polyethylene glycol (PEG 3350)	0.3–1.5%	0.1–0.4 g/L
Polyvinylpyrrolidone 40 (PVP40)		0.05–4%
Alpha-cyclodextrin	8–40 mM	
Beta-cyclodextrin	1–5 mM	
Putrescine, spermidine, and spermine	0.1 M	
Formamide	0.1%	
Reducing agents		
β-Mercaptoethanol (BME)	2–5 mM	1–10 mM
Dithiothreitol (DTT)	1 mM	0.1–10 mM
Tris(2-carboxyethyl)phosphine (TCEP)	1–5 mM	1–50 mM

Size exclusion chromatography (SEC), despite its limitations, is the most accepted and reliable technique to detect correctly oligomerized proteins. SEC can be coupled in-line to a light-scattering device (SEC-MALS) to measure the absolute molar mass, size, and shape of macromolecules in solution. One major disadvantage is that it is time consuming, although this can be partially circum-

vented by the use of mini-analytical columns [66]. Other drawbacks are solubilization of reversible aggregates as a result of dilution effects, and the loss of larger aggregates in the pre-column filters [67].

A popular method in crystallography or NMR studies is the fluorescence-based thermal shift (ThermoFluor) assay. It uses an environmentally sensitive dye, Sypro Orange, to monitor the thermal stability of a pure protein under different buffer conditions [68,69]. Real-time PCR machines with a fluorescent detector are used to compare melting curve shifts in Tm (the midpoint of the unfolding transition). However, this approach has some disadvantages: it cannot provide information regarding the oligomeric state of the protein, it cannot be used in the presence of additives such as detergents, and the presence of intrinsic fluorescent aggregates makes it difficult to interpret. Moreover, the ability to predict best crystallization conditions is still under debate within the protein-producing community.

Choosing the analysis approach should take into account the large number of tested variables (different buffers, pH, additives, salt, etc.) and the fact that no single method is optimal for identifying all types of aggregates. Unfortunately, approaches that provide the most accurate information regarding the oligomeric state of the protein are not applicable in high-throughput screening (HTS).

Our approach is to combine a fast standard SDS–PAGE analysis that allows screening of many variables, with analytic SEC, that provides information regarding the oligomeric conformation of the protein. SDS-PAGE allows the easy selection of the most promising condition and considerably eliminates the number of samples to be analyzed by SEC.

3.4. Our comprehensive approach for optimizing solubility during purification

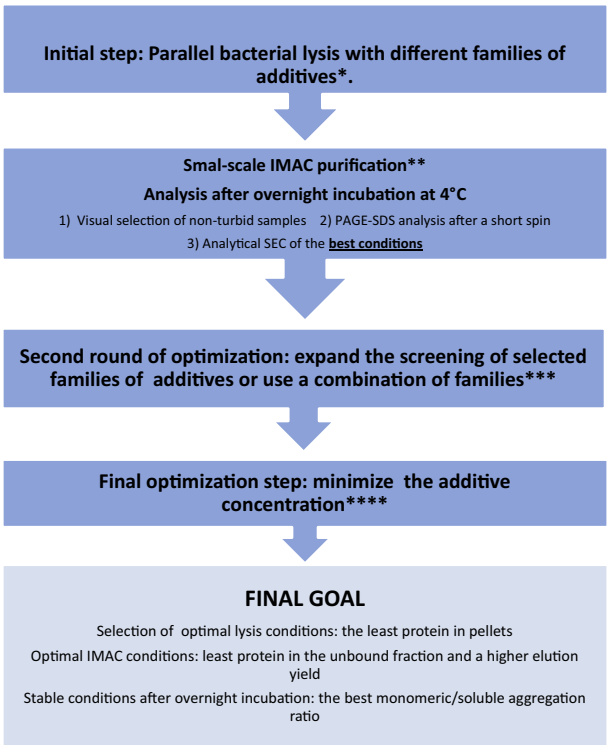
In standard procedures, purified or partially purified protein is used to screen the solubility conditions. We have found that solubility problems at the purification level should be tackled as early as the cell-lysate step. Starting the screen at this initial step, followed by analyzing the oligomeric state of partially purified protein at different purification steps provides maximum information on solubility issues and improves the final output, since it may rescue the protein fraction that was partially insoluble in the bacteria and that was mistakenly considered as inclusion bodies [63].

Our screening consists of a short list of solubility-promoting additives during cell lysis (Flow chart 2), followed by a quick parallel capture step of small-scale IMAC purification (or any other capture method) in the presence of selected additives. Samples are then analyzed on PAGE–SDS and include insoluble lysis extracts, unbound fractions, and eluted protein. Incubation of the eluted protein at 4 °C for 16 h will allow the detection of slow-forming aggregates. Samples that do not contain aggregates (based on the initial visual detection of turbidity and then based on PAGE–SDS analysis) will be selected for further testing on analytical SEC, searching for the best monomer/soluble aggregate ratio.

From this first screening, we obtain valuable information regarding: (a) the effect of representative additive groups on optimal lysis, (b) better binding to the capture resin, and (c) the best yield of native oligomeric conformation over time.

A short alternative method is to split a common eluate after IMAC purification, and dilute it in buffers with different additives. Its drawback is that in this way only data regarding protein stability are obtained, and important information on lysis yield and resin capture is lost.

We therefore prefer to check, using this first screen, a non-ionic and a zwiterionic detergent, a mild chaotrope, an osmolyte, and



Flow chart 2. Screening methodology showing the best additives for purifying prone-to-aggregate proteins. *Different families of additives were used to stabilize folding and prevent aggregation during purification, as can be seen in Table 3. As a start screen we added additives to a 50 mM buffer (HEPES, TrisHCl, etc. pH 7.5–8.0) with a relatively high NaCl concentration (0.5 M) and reducing agents according to the presence of free Cys residues (see Table 2 and 3). • Control. • Non-ionic detergent (0.5%). • Zwittergent (0.5%). • Osmolyte (0.5 M). • Urea or GuHCl (1 M). • Add Arg only in the elution buffer (0.5 M). **_Alternative strategy:_ Split purified protein after IMAC purification and dilute in buffers with different additives. ***Use the same protocol as in first round. Check similar or a combination (synergism) of additives around the best condition obtained from the first step (see Table 3). ****Same protocol, using lower additive concentrations of the best results from previous rounds. _Alternative:_ Additive concentrations can be drastically reduced in the later purification steps without changing the additive concentration during lysis and the capture step.

Arginine (only in the elution buffer), all of which are in the same relatively high salt buffer. Deliberately, we prefer not to introduce pH as a variable in this first screening because: (1) IMAC binding itself constrains pH options, (2) we try to reduce as much as possible the number of variables to check, and we consider that pH is less important at this point, and (3) there is indirect evidence from refolding screens that buffers with pH far away from the protein's pI, are best for protein refolding [70]. We prefer to use pH screenings at later purification stages and storage conditions.

Next, the optimization rounds check whether additives from the same family yield even better results (Table 3). This step can be improved by combining additives from different families in order to obtain a synergistic effect. In the final round, the additives' concentration is optimized. This last optimization is very important when additive concentration use in the capture step is not compatible with the next steps of purification or with protein application. We employ this strategy in most of the projects that involve protein production in bacteria, with high success rates in dozens of cases. A few successful examples can be found in the following references [25–27,57–59].

Optionally, a drastic dilution of additive concentration at later stages of purification can be considered, since the beneficial effects of the chosen additives are most significant at the cell lysis and early purification steps, and the additive concentration can be considerably reduced in the later purification steps [63].

We wish to emphasize that the initial buffer used during the cell-lysis stage and capture steps does not necessarily match the protein's requirements for ensuring the next chromatographic steps, the optimal storage conditions, or for protein concentrations for structural studies and other downstream applications.

4. Conclusions

In an academic research environment it is very tempting to avoid extensive and laborious screening when dealing with protein solubility problems. HTS platforms that allow parallel screening of multiple parameters are not commonly available in most laboratories. Many researchers use available vectors and strains in their surroundings, adopt standard buffer conditions at the purification step and end up dealing with aggregation problems at the end of the whole production process. Moreover, individual parameters in the process are frequently being tested in a linear process, changing a single parameter at a time. This time-consuming, and often frustrating approach can lead to a dead end due to the formation of protein aggregates. In this review we described a variety of expression and purification tools used to increase the efficiency of producing soluble proteins. Most importantly, we tried to provide a simple and modular toolbox for designing a minimized expression and purification screen. We believe that following these hierarchic rules will make soluble protein production in *E. coli*-based systems more efficient, simpler, and less costly. Moreover, the rules for designing this screen can be applied to the HTS formats, as well.

In the 'protein expression' section we focused on the pros and cons of the initial cloning steps, the possible fusion partners, suitable bacterial hosts, and the optimal induction conditions. The most beneficial approaches were then combined to form a minimized hierarchical screening platform, where the main goal is to obtain maximal yield of the correct oligomeric conformation along with the minimal presence of insoluble proteins, or soluble aggregates, in the shortest possible time. Although obtaining soluble proteins during the expression steps is an essential prerequisite for the subsequent purification steps, aggregation can also occur at later stages, e.g., during the purification process. In the 'protein purification' section we describe in detail: (1) the factors that can affect solubility during the purification step, (2) general and environmental considerations that can affect protein purification, and (3) various methods for monitoring protein aggregation and their limitations. Finally, we describe our approach for screening the best additives and the conditions needed for purifying prone-to-aggregate proteins.

Despite the constant flow of new and improved vectors, host strains, HTS platforms, various buffer formulations, additives and other factors, the experience accumulated at our core facilities indicates that combining traditional strategies for screening parallel factors, both in the expression and the purification steps, results in a synergistic effect, thus allowing increased production efficacy of soluble proteins using the simple tools readily available in most research laboratories. We believe that our integrated and hierarchical approach follows the "20/80 rule", achieving 80% of the benefits with only 20% of the resources, while addressing the vast majority of solubility problems. Finally, we believe that this approach simplifies the decision whether to continue to invest in *E. coli* production or instead to explore the potential of using other hosts.

References

[1] Philo, J.S. and Arakawa, T. (2009) Mechanisms of protein aggregation. Curr. Pharm. Biotechnol. 10, 348–351.

[2] Morris, A.M., Watzky, M.A. and Finke, R.G. (2009) Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. Biochim. Biophys. Acta 1794, 375–397.

[3] Lee, H.J., McAuley, A., Schilke, K.F. and McGuire, J. (2011) Molecular origins of surfactant-mediated stabilization of protein drugs. Adv. Drug Deliv. Rev. 63, 1160–1171.

[4] Peng, Z., Oldfield, C.J., Xue, B., Mizianty, M.J., Dunker, A.K., Kurgan, L. and Uversky, V.N. (2013) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome Cell. Mol. Life Sci., http://dx.doi.org/10.1007/s00018-013-1446-6.

[5] Uversky, V.N. (2013) Unusual biophysics of intrinsically disordered proteins. Biochim. Biophys. Acta 1834, 932–951.

[6] Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113–130.

[7] Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130.

[8] Qing, G., Ma, L.C., Khorchid, A., Swapna, G.V., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G.T., Ikura, M. and Inouye, M. (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. Nat. Biotechnol. 22, 877–882.

[9] Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T. and Valla, S. (2013) A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. Microb. Cell Fact. 12, 26.

[10] Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 72, 211–222.

[11] Esposito, D. and Chatterjee, D.K. (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr. Opin. Biotechnol. 17, 353–358.

[12] Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 60, 523–533.

[13] Sorensen, H.P., Sperling-Petersen, H.U. and Mortensen, K.K. (2003) A favorable solubility partner for the recombinant expression of streptavidin. Protein Expr. Purif. 32, 252–259.

[14] Kapust, R.B. and Waugh, D.S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8, 1668–1674.

[15] LaVallie, E.R., Lu, Z., Diblasio-Smith, E.A., Collins-Racie, L.A. and McCoy, J.M. (2000) Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. Methods Enzymol. 326, 322–340.

[16] Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol. Bioeng. 65, 382–388.

[17] Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D. and Butt, T.R. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J. Struct. Funct. Genomics 5, 75–86.

[18] Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X. and Butt, T.R. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci. 15, 182–189.

[19] Zou, Z., Cao, L., Zhou, P., Su, Y., Sun, Y. and Li, W. (2008) Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. J. Biotechnol. 135, 333–339.

[20] Raran-Kurussi, S. and Waugh, D.S. (2012) The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. PLoS ONE 7, e49589.

[21] Lebendiker, M. and Danieli, T. (2011) Purification of proteins fused to maltose-binding protein. Methods Mol. Biol. 681, 281–293.

[22] Nallamsetty, S. and Waugh, D.S. (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expr. Purif. 45, 175–182.

[23] McCoy, J., and La Ville, E. (2001) Expression and purification of thioredoxin fusion proteins, in: John E. Coligan, et al. (Eds.), Current protocols in protein science, (Chapter 6, Unit 6 7).

[24] Packman, L.C., Borges, A. and Perham, R.N. (1988) Amino acid sequence analysis of the lipoyl and peripheral subunit-binding domains in the lipoate acetyltransferase component of the pyruvate dehydrogenase complex from Bacillus stearothermophilus. Biochem. J. 252, 79–86.

[25] Amartely, H., David, A., Lebendiker, M., Benyamini, H., Izraeli, S. and Friedler, A. (2013) The STIL protein contains intrinsically disordered regions that mediate its protein–protein interactions Chem. Commun., http://dx.doi.org/10.1039/c3cc45096a.

[26] Katz, C., Benyamini, H., Rotem, S., Lebendiker, M., Danieli, T., Iosub, A., Refaely, H., Dines, M., Bronner, V., Bravman, T., Shalev, D.E., Rudiger, S. and Friedler, A. (2008) Molecular basis of the interaction between the antiapoptotic Bcl-2 family proteins and the proapoptotic protein ASPP2. Proc. Natl. Acad. Sci. U.S.A. 105, 12277–12282.

[27] Rotem, S., Katz, C., Benyamini, H., Lebendiker, M., Veprintsev, D., Rudiger, S., Danieli, T. and Friedler, A. (2008) The structure and interactions of the proline-rich domain of ASPP2. J. Biol. Chem. 283, 18990–18999.

[28] Song, J.A., Lee, D.S., Park, J.S., Han, K.Y. and Lee, J. (2012) The N-domain of *Escherichia coli* phosphoglycerate kinase is a novel fusion partner to express aggregation-prone heterologous proteins. Biotechnol. Bioeng. 109, 325–335.

[29] Chatterjee, D.K. and Esposito, D. (2006) Enhanced soluble protein expression using two new fusion tags. Protein Expr. Purif. 46, 122–129.

[30] Segatori, L., Paukstelis, P.J., Gilbert, H.F. and Georgiou, G. (2004) Engineered DsbC chimeras catalyze both protein oxidation and disulfide-bond

isomerization in *Escherichia coli*: Reconciling two competing pathways. Proc. Natl. Acad. Sci. U.S.A. 101, 10018–10023.
[31] Zhang, Z., Li, Z.H., Wang, F., Fang, M., Yin, C.C., Zhou, Z.Y., Lin, Q. and Huang, H.L. (2002) Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in *Escherichia coli*. Protein Expr. Purif. 26, 218–228.
[32] Waugh, D.S. (2011) An overview of enzymatic reagents for the removal of affinity tags. Protein Expr. Purif. 80, 283–293.
[33] Arnau, J., Lauritzen, C., Petersen, G.E. and Pedersen, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr. Purif. 48, 1–13.
[34] Vergis, J.M. and Wiener, M.C. (2011) The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. Protein Expr. Purif. 78, 139–142.
[35] Sun, C., Liang, J., Shi, R., Gao, X., Zhang, R., Hong, F., Yuan, Q. and Wang, S. (2012) Tobacco etch virus protease retains its activity in various buffers and in the presence of diverse additives. Protein Expr. Purif. 82, 226–231.
[36] Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U. and Gileadi, O. (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expr. Purif. 59, 94–102.
[37] Makhoul, C.H. and Trifonov, E.N. (2002) Distribution of rare triplets along mRNA and their relation to protein folding. J. Biomol. Struct. Dyn. 20, 413–420.
[38] Dyson, M.R. (2010) Selection of soluble protein expression constructs: the experimental determination of protein domain boundaries. Biochem. Soc. Trans. 38, 908–913.
[39] Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M.J. and Walker, J.E. (2000) Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F(1)F(o) ATP synthase. FEBS Lett. 482, 215–219.
[40] Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Hogbom, M., van Wijk, K.J., Slotboom, D.J., Persson, J.O. and de Gier, J.W. (2008) Tuning *Escherichia coli* for membrane protein overexpression. Proc. Natl. Acad. Sci. U.S.A. 105, 14371–14376.
[41] Hjelm, A., Schlegel, S., Baumgarten, T., Klepsch, M., Wickstrom, D., Drew, D. and de Gier, J.W. (2013) Optimizing E. coli-based membrane protein production using Lemo21(DE3) and GFP-fusions. Methods Mol. Biol. 1033, 381–400.
[42] Schlegel, S., Rujas, E., Ytterberg, A.J., Zubarev, R.A., Luirink, J. and de Gier, J.W. (2013) Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels. Microb. Cell Fact. 12, 24.
[43] Prinz, W.A., Aslund, F., Holmgren, A. and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. J. Biol. Chem. 272, 15661–15667.
[44] Levy, R., Weiss, R., Chen, G., Iverson, B.L. and Georgiou, G. (2001) Production of correctly folded Fab antibody fragment in the cytoplasm of *Escherichia coli* trxB gor mutants via the coexpression of molecular chaperones. Protein Expr. Purif. 23, 338–347.
[45] Berndt, C., Lillig, C.H. and Holmgren, A. (2008) Thioredoxins and glutaredoxins as facilitators of protein folding. Biochim. Biophys. Acta 1783, 641–650.
[46] Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P. and Berkmen, M. (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb. Cell Fact. 11, 56.
[47] Thomas, J.G. and Baneyx, F. (1996) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing Heat-shock proteins. J. Biol. Chem. 271, 11141–11147.
[48] Maier, T., Ferbitz, L., Deuerling, E. and Ban, N. (2005) A cradle for new proteins: trigger factor at the ribosome. Curr. Opin. Struct. Biol. 15, 204–212.
[49] de Marco, A., Vigh, L., Diamant, S. and Goloubinoff, P. (2005) Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. Cell stress & chaperones. 10, 329–339.
[50] Diamant, S., Eliahu, N., Rosenthal, D. and Goloubinoff, P. (2001) Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses. J. Biol. Chem. 276, 39586–39591.
[51] Singh, L.R., Chen, X., Kozich, V. and Kruger, W.D. (2007) Chemical chaperone rescue of mutant human cystathionine beta-synthase. Mol. Genet. Metab. 91, 335–342.
[52] Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. 41, 207–234.
[53] Schlapschy, M. and Skerra, A. (2011) Periplasmic chaperones used to enhance functional secretion of proteins in E. coli. Methods Mol. Biol. 705, 211–224.
[54] Liddy, N., Molloy, P.E., Bennett, A.D., Boulter, J.M., Jakobsen, B.K. and Li, Y. (2010) Production of a soluble disulfide bond-linked TCR in the cytoplasm of *Escherichia coli* trxB gor mutants. Mol. Biotechnol. 45, 140–149.
[55] de Marco, A. and De Marco, V. (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. J. Biotechnol. 109, 45–52.
[56] Kolaj, O., Spada, S., Robin, S. and Wall, J.G. (2009) Use of folding modulators to improve heterologous protein production in *Escherichia coli*. Microb. Cell Fact. 8, 9.
[57] Gabizon, R., Brandt, T., Sukenik, S., Lahav, N., Lebendiker, M., Shalev, D.E., Veprintsev, D. and Friedler, A. (2012) Specific recognition of p53 tetramers by peptides derived from p53 interacting proteins. PLoS ONE 7, e38060.
[58] Reingewertz, T.H., Benyamini, H., Lebendiker, M., Shalev, D.E. and Friedler, A. (2009) The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. Protein engineering, design & selection: PEDS. 22, 281–287.
[59] Siman, P., Blatt, O., Moyal, T., Danieli, T., Lebendiker, M., Lashuel, H.A., Friedler, A. and Brik, A. (2011) Chemical synthesis and expression of the HIV-1 Rev protein. Chembiochem 12, 1097–1104.
[60] Cromwell, M.E., Hilario, E. and Jacobson, F. (2006) Protein aggregation and bioprocessing. AAPS J. 8, E572–E579.
[61] Churion, K.A. and Bondos, S.E. (2012) Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification. Methods Mol. Biol. 896, 415–427.
[62] Lebendiker, M. M., M. and Friedler, A. (2013) A screening methodology for purifying proteins with aggregation problems, Methods Mol. Biol. (Accepted for publication).
[63] Leibly, D.J., Nguyen, T.N., Kao, L.T., Hewitt, S.N., Barrett, L.K. and Van Voorhis, W.C. (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. PLoS ONE 7, e52482.
[64] Bondos, S.E. and Bicknell, A. (2003) Detection and prevention of protein aggregation before, during, and after purification. Anal. Biochem. 316, 223–231.
[65] Gileadi, O. (2012) Rational optimization of protein stability, P4EU Workshop on Protein Purification, Barcelona.
[66] Sala, E. and 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.
[67] Carpenter, J.F., Randolph, T.W., Jiskoot, W., Crommelin, D.J., Middaugh, C.R. and Winter, G. (2010) Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: essential need to use orthogonal methods to assure the quality of therapeutic protein products. J. Pharm. Sci. 99, 2200–2208.
[68] Niesen, F.H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212–2221.
[69] Nettleship, J.E., Brown, J., Groves, M.R. and Geerlof, A. (2008) Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. Methods Mol. Biol. 426, 299–318.
[70] Coutard, B., Danchin, E.G., Oubelaid, R., Canard, B. and Bignon, C. (2012) Single pH buffer refolding screen for protein from inclusion bodies. Protein Expr. Purif. 82, 352–359.
[71] Hamada, H., Arakawa, T. and Shiraki, K. (2009) Effect of additives on protein aggregation. Curr. Pharm. Biotechnol. 10, 400–407.
[72] Chi, E.Y., Krishnan, S., Randolph, T.W. and Carpenter, J.F. (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm. Res. 20, 1325–1336.
[73] Ohtake, S., Kita, Y. and Arakawa, T. (2011) Interactions of formulation excipients with proteins in solution and in the dried state. Adv. Drug Deliv. Rev. 63, 1053–1073.
[74] Street, T.O., Bolen, D.W. and Rose, G.D. (2006) A molecular mechanism for osmolyte-induced protein stability. Proc. Natl. Acad. Sci. U.S.A. 103, 13997–14002.
[75] Timasheff, S.N. (1998) Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. Adv. Protein Chem. 51, 355–432.
[76] Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y. and Timasheff, S.N. (2007) Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. Biophys. Chem. 127, 1–8.
[77] Rudolph, R. a. F., S. (1990) Process for obtaining renaturated proteins.
[78] Umetsu, M., Tsumoto, K., Hara, M., Ashish, K., Goda, S., Adschiri, T. and Kumagai, I. (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, *The Journal of biological chemistry*. 278, 8979–8987.
[79] Russo, D. (2008) The impact of kosmotropes and chaotropes on bulk and hydration shell water dynamics in a model peptide solution. Chem. Phys. 345, 200–211.
[80] De Bernardez Clark, E., Schwarz, E. and Rudolph, R. (1999) Inhibition of aggregation side reactions during in vitro protein folding. Methods Enzymol. 309, 217–236.
[81] Golovanov, A.P., Hautbergue, G.M., Wilson, S.A. and Lian, L.Y. (2004) A simple method for improving protein solubility and long-term stability. J. Am. Chem. Soc. 126, 8933–8939.