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,