Available online at www.sciencedirect.com
ScienceDirect
Protein Expression and Purification 55 (2007) 209–224
Protein Expression & Purification
www.elsevier.com/locate/yprep

Review

Confronting high-throughput protein refolding using high pressure and solution screens

M. Walid Qoronfleh [a,*], Lyndal K. Hesterberg [b], Matthew B. Seefeldt [b]

[a] University of Michigan and Core Technology Alliance-CTA, 1024 Wolverine Tower, 3003 State Street, Ann Arbor, MI 48109-1274, USA
[b] BaroFold Inc., 1825 33rd Street, Boulder, CO 80301, USA

Received 29 November 2006, and in revised form 7 May 2007
Available online 2 June 2007

Abstract

Over-expression of heterologous proteins in *Escherichia coli* is commonly hindered by the formation of inclusion bodies. Nevertheless, refolding of proteins *in vitro* has become an essential requirement in the development of structural genomics (proteomics) and as a means of recovering functional proteins from inclusion bodies. Many distinct methods for protein refolding are now in use. However, regardless of method used, developing a reliable protein refolding protocol still requires significant optimization through trial and error. Many proteins fall into the category of "Challenging" or "Difficult to Express" and are problematic to refold using traditional chaotrope-based refolding techniques. This review discusses new methods for improving protein refolding, such as implementing high hydrostatic pressure, using small molecule additives to enhance traditional protein refolding strategies, as well as developing practical methods for performing refolding studies to maximize their reliability and utility. The strategies examined here focus on high-throughput, automated refolding screens, which can be applied to structural genomic projects.

Keywords: Inclusion bodies; Aggregation; Misfolding; Solubilization; Protein recovery; Protein refolding; Small molecules assisted refolding; Fractional factorial screen; High hydrostatic pressure; Automation; High-throughput; Structural genomics

The expression of proteins in transformed microorganisms has been one of the fundamental techniques in the development and expansion of modern biological research. Numerous expression systems are available, allowing both high and low levels of protein expression in a variety of prokaryotic or eukaryotic organisms [1]. Despite the successes and growth of expression technology, significant pitfalls still exist. Expression of recombinant proteins often results in the accumulation of inactive and improperly folded proteins in the form of aggregates and occurs commonly when eukaryotic, disulfide containing, post translational modified, or multimeric proteins are expressed in insect and bacterial systems [2,3]. While some expression systems lead to high level of soluble aggregates, expression in *Escherichia coli* often leads to insoluble aggregates known as inclusion bodies. Inclusion bodies are dense structures of the misfolded expressed polypeptides that arise from the inability of cellular machinery to process and refold the polypeptide correctly [3]. In some cases, inclusion body formation is a result from the reducing environment of the cytoplasm of *E. coli* preventing native disulfide bond formation. Additionally, the cellular machinery of *E. coli* is incapable of handling the high levels of expression that occur during typical recombinant protein production and results in inclusion body formation. Optimization of soluble protein expression is often the strategy of choice when trying to obtain bioactive protein [1,4]. Numerous labor-intensive expression systems and culture conditions have been developed that attempt to prevent inclusion body formation. However, the formation of inclusion bodies is still common and often unavoidable.

* Corresponding author. Present address: 3223 Dunwoodie Rd., Ann Arbor, MI 48105, USA. Fax: +1 734 764 8510.
E-mail address: walidq@umich.edu (M.W. Qoronfleh).

1046-5928/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2007.05.014

Although inclusion bodies are often considered undesirable, their formation can be advantageous. The benefits associated with inclusion bodies include: (i) high level, inexpensive, expression and homogeneity of the target protein, (ii) proteolytic resistance, and as a result, lower levels of degradation of the expressed protein, and (iii) easy isolation and purification of the inclusion bodies from host cells [5]. These benefits can only be achieved if the protein of interest can be refolded to obtain native protein with high yields. Fortunately, inclusion bodies can be solubilized and refolded to release the misfolded or aggregated protein by using high hydrostatic pressure or chaotropes. Unfortunately, there is no universal refolding method or buffer. A variety of methods may be used to refold inclusion bodies; however, the method and refolding solution conditions can greatly impact the subsequent refolding step and the cost of the overall process [6,7]. While protein refolding is often a strategy of last resort due to unpredictability, time requirement, and operational issues of the renaturation process, the literature demonstrates the applicability and success of protein refolding techniques [6]. This review summarizes the current state-of-the-art in high-throughput protein refolding, describing refolding screens that rapidly identify initial conditions that successfully result in folded protein. Additionally, new protein refolding techniques, such as high pressure refolding, have promise for improving refolding yields in many protein classes [8].

Misfolding and aggregation

The structure of proteins is maintained by a delicate thermodynamic balance of hydrogen bonding, and hydrophobic and electrostatic interactions. These interactions are weak relative to covalent bonds and proteins are inherently unstable and susceptible to the formation of non-native aggregates and precipitates. Non-native protein aggregation (described hereafter as ''aggregation'') describes the assembly of native or structurally perturbed proteins to aggregates containing non-native protein structures [9]. Aggregation occurs commonly *in vivo* and living cells have developed sophisticated mechanism to either prevent aggregation (i.e. molecular chaperones) or conduct housekeeping to degrade misfolded proteins [10]. During *in vitro* protein processing, aggregation is often irreversible at physiological conditions and may result in the formation of high levels of non-native, intermolecular β-sheet structures [11].

The reader is referred to an excellent review on protein folding, misfolding and aggregation in pharmaceutical proteins by Chi et al. [12]. Briefly, protein aggregation proceeds through specific pathways that are initiated by instability of the native protein conformation or colloid instability associated with protein–protein interactions. Conditions such as temperature, solution pH, ligands and cosolutes, salt type and concentration, preservatives, and surfactants all modulate protein structure and protein–protein interactions, and thus aggregation propensity. For aggregates that form from native protein instability, it appears that aggregates may form from protein structures present within the native state that demonstrate an expanded conformation and are often the result of non-specific hydrophobic interactions [13,14]. Consequently, aggregation is controlled by the conformational stability of the native protein relative to that of the aggregation transition state. Recently, it has been reported that proteins can form aggregates due to colloidal instability, even in solution conditions which thermodynamically greatly favor the native conformation [15]. These molecular assembly reactions are a result of intermolecular attractions. For example, GCSF at pH 7.0 has been demonstrated to have a large $\Delta G_{unfolding,}$ yet the protein aggregates readily due to colloidal instability arising from attractive electrostatic interactions [15]. Due to the myriad of aggregation mechanisms in all proteins, it is not surprising that protein aggregation is a widespread problem in all aspects of protein processing, both *in vivo* and *in vitro*. It is therefore imperative that any group wishing to refold significant numbers of proteins have access to state of the art technology and methodology.

During the process of refolding, solubilization of the protein of interest can readily be achieved by the addition of high concentrations of chaotropes, such as urea or GdHCl. The complications arise when the chaotrope is removed. The current folding theories propose that protein folding occurs along a ''free-energy funnel'', where the protein forms structure along a landscape of intermediates of lower free energy [16,17] (Fig. 1). As the protein navigates

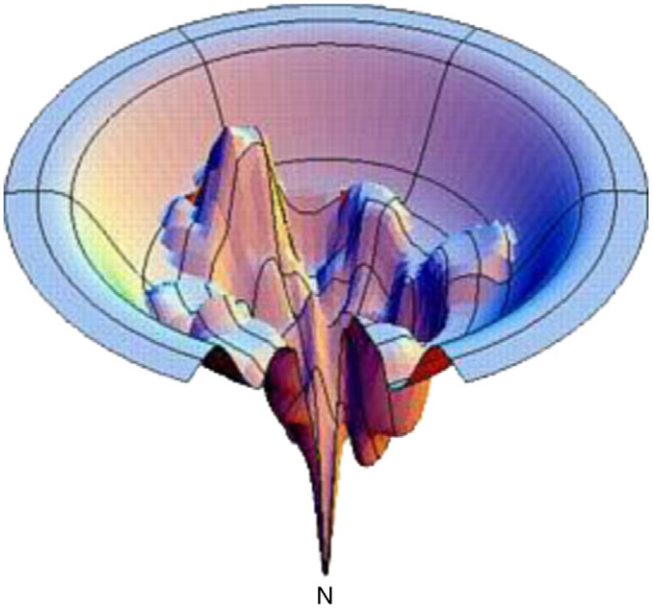

Fig. 1. Folding landscape of a protein, as depicted by Dill et al. [18]. The upper rim of the funnel is representative of high energy, denatured protein. As the thermodynamic energy state of the protein is decreased, the protein begins to fold as it migrates down the energy funnel. However, kinetic traps are often encountered which prevent the formation of native structure and can lead to reaggregation since the folding intermediates are often aggregation prone.