Journal of Structural and Functional Genomics **6:** 177–182, 2005.
DOI 10.1007/s10969-005-2827-3

On-column protein refolding for crystallization

Natalia Oganesyan[1], Sung-Hou Kim[1,2] & Rosalind Kim[1,*]
[1]*Berkeley Structural Genomics Center, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA;* [2]*Department of Chemistry, University of California, Berkeley, California, 94720, USA;* **Author for correspondence (e-mail: R-Kim@lbl.gov; fax: +1-510-486-5272)*

Received 31 July 2004; accepted in revised form 15 February 2005

Key words: crystallization, recombinant proteins, refolding

Abstract

One major bottleneck in protein production in *Escherichia coli* for structural genomics projects is the formation of insoluble protein aggregates (inclusion bodies). The efficient refolding of proteins from inclusion bodies is becoming an important tool that can provide soluble native proteins for structural and functional studies. Here we report an on-column refolding method established at the Berkeley Structural Genomics Center (BSGC). Our method is a combination of an 'artificial chaperone-assisted refolding' method previously proposed and affinity chromatography to take advantage of a chromatographic step: less time-consuming, no filtration or concentration, with the additional benefit of protein purification. It can be easily automated and formatted for high-throughput process.

Introduction

In the past several years, more than 25 structural genomics projects from different countries have been organized [1]. They generate a vast amount of information about protein properties on each step of the structural genomics pipeline, from cloning, expression, and purification to structure determination. Collected data leads to understanding of protein structure, function and regulation, and reveals a global view of the protein structure universe [2].

Production of soluble proteins plays a critical role in high-throughput methods used in structural genomics. Due to fast growth, easy handling and low cost, *E. coli* has been the principal expression system of choice for most structural genomics projects. However, often recombinant proteins overproduced in *E. coli* accumulate as insoluble aggregates known as inclusion bodies (IB). IB formation can be minimized, or potentially avoided, by applying complex efforts to enhance production of soluble protein: cloning and expression of homologous genes, variations in expression vectors, host strains and growth conditions. But protein production from inclusion bodies has a number of merits: they are produced in high yields, even those that are toxic for bacterial cells; are generally protected from proteolytic degradation, and can be easily solubilized. The only challenge is to convert inclusion bodies to properly folded, biologically active proteins. Conventional methods for refolding insoluble recombinant proteins include slow dialysis or dilution of urea- or guanidinium-HCl-solubilized IB into refolding buffer, or chromatographic refolding using packed columns. Chromatographic methods can include solvent-exchange size exclusion chromatography and immobilization of the denatured protein onto a matrix and subsequent denaturant dilution to promote refolding. A vast amount of data in the literature provides information aimed at enhancing the refolding yield of inclusion body proteins by adding certain low molecular weight additives to reduce protein aggregation. Surfactants and detergents have proven to be very efficient folding aids, and have been shown to work with a

variety of proteins in dilution or chromatographic refolding [3–6]. One drawback in the use of surfactants and detergents is that they are difficult to remove, a direct result of their ability to bind to proteins and to form micelles. Rozema and Gellman [7] developed a new dilution-based folding strategy in which the denatured protein is first exposed to a detergent-containing solution to prevent aggregation, followed by stripping of the detergent with cyclodextrin to promote refolding. This method has been claimed to mimic the GroEL–GroES chaperonin action *in vivo* and has been named 'artificial chaperone-assisted refolding'. By this method, several proteins have been successfully refolded [8–10]. Although their method is attractive for its efficacy and practicality, it is time-consuming and accompanied by loss of protein during filtration and concentration of large volumes. Here we describe an on-column chemical refolding method that we use at BSGC for insoluble His-tagged proteins expressed in *E. coli*. IB solubilized in urea are first bound to an affinity column and exposed to a detergent wash to prevent misfolding. This is followed by a β-cyclodextrin wash that removes the detergent and promotes correct folding. The target protein is eluted with imidazole, goes through further purification steps: ion-exchange (IEX) and/or size exclusion chromatography (SEC), and is evaluated by dynamic light scattering (DLS), analytical size exclusion chromatography and/or by circular dichroism (CD) spectroscopy. We have been able to obtain 30–100% refolding in 7 out of 10 tested proteins. Six of the seven refolded proteins were able to produce crystals of varying qualities.

Materials and methods

Recombinant plasmids and expression hosts

The modified ligation-independent cloning (LIC) system [11] was used to clone genes encoding proteins from *Mycoplasma pneumoniae* and *M. genitalium*, and their homologues from other species. The amplified PCR product was prepared for vector insertion by purification, quantitation and treatment with T4 DNA polymerase (New England Biolabs, Beverley, MA) in the presence of 1 mM dTTP. The prepared insert was annealed into the LIC expression vector pB3, a derivative of pET21a (Novagen, Madison, WI) that expresses the cloned gene fused with an N-terminal 6-His-TEV (tobacco etch virus protease) cleavage sequence, and transformed into chemical competent DH5α cells to obtain fusion clones. Protein was expressed in *E. coli* strain BL21(DE3) Star/pSJS1244 [12].

Isolation of inclusion bodies

For our study, we chose 10 proteins that were expressed as inclusion bodies (BSGC website: http://www.strgen.org). The cells transformed with the constructs were grown in 0.5 l of auto-inducing growth media (Dr. William Studier, Brookhaven National Laboratory, personal communication) with carbenicillin (100 μg/ml) and spectinomycin (30 μg/ml). After 24 h of incubation at 37 °C at 250 rpm, the cultures were collected by centrifugation, resuspended in 50 mM HEPES, pH 7.0, 100 mM NaCl, treated with lysozyme at a final concentration of 0.1 mg/ml, and opened with a microfluidizer (MFIC, Newton, MA). After centrifugation at 10,000 g for 20 min, the pellet was washed with 1 M urea, 2% Triton X-100 and centrifuged again. The pellet was then solubilized in 30 ml of denaturing buffer containing 20 mM Tris–HCl, pH 8.0, 8 M urea, stirred for 2 h at room temperature, and centrifuged.

In vitro *refolding*

Solubilized IB were bound to Ni-NTA resin (Qiagen, Valencia, CA) pre-equilibrated in denaturing buffer by batch-absorption overnight at room temperature. The amount of resin used was calculated according to manufacturer's recommendation (5 mg of protein/1 ml of resin). On-column renaturation and purification were performed the next morning by several changes of buffers. The resin was packed into a glass Econo-Column of 2×5 cm diameter (Bio-Rad, Hercules, CA). All chromatography steps were done under gravity. First, the column was washed using the denaturing buffer containing 20 mM imidazole to remove nonspecifically bound contaminants. β-mercaptoethanol at a final concentration of 10 mM was added to this buffer if the target protein contained cysteines.

All renaturation steps were carried out in Buffer A (20 mM Tris–HCl, pH 7.5, 0.1 M NaCl). The pH of buffer A should be at least 1.0 pH units away from the pI of the protein to avoid protein precipitation. In the next step, the column was washed with 10 column volumes (CV) of 20 mM Tris–HCl, pH 7.5, containing 0.1% Triton X-100 (Anatrace, Maumee, OH) and 500 mM NaCl. This was followed by a wash with 10 CV of Buffer A containing 5 mM β-cyclodextrin (Sigma, St. Louis, MO) to remove detergent from the protein–detergent complex and to allow the protein to refold. An additional wash with buffer A was applied to remove remaining β-cyclodextrin before elution. Refolded protein was eluted with Buffer A supplemented with 300 mM imidazole. The eluted fractions containing soluble refolded protein were run on a sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) and subjected to further purification steps (IEX and/or SEC).

Protein folding assay

The refolded proteins were analyzed by DLS to determine monodispersity and analytical SEC to determine molecular weight. DLS measurements were carried out with the DynaPro 99 instrument (Wyatt Technology, Santa Barbara, CA) at a protein concentration of 1 mg/ml in 20 mM Tris–HCl, pH 7.5, and 100 mM NaCl. Refolded protein was loaded on a G4000SWXL column (TosoHaas, Montgomeryville, PA) and eluted at 0.5 ml/min in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA. To determine secondary structure of the refolded protein, a CD spectra of refolded proteins was analyzed. CD spectra of protein samples (0.4 mg/ml) in 1 mm path length quartz cuvette was recorded as an average of four scans with a Jasco J-810 spectropolarimeter (Jasco Inc., Easton, MD) over a range of 190–260 nm on a millidegree ellipticity scale.

Results and discussion

To obtain pure soluble protein from inclusion bodies for crystallographic purposes, we developed a new refolding method by using a column-based approach with detergent and cyclodextrin as low molecular weight additives. Refolding in the presence of a detergent followed by addition of cyclodextrin had been proposed by Rozema and Gellman [7]. Their method utilized a dilution approach, whereby a protein denatured by urea or guanidinium-HCl was diluted to low concentration in a detergent-containing buffer, then diluted again in the presence of a cyclodextrin solution. After refolding was completed, the protein solution had to be passed through a 0.22 μm filter to remove aggregated protein, and then through a 10 kDa cut-off filter to remove detergent and cyclodextrin. In order to adapt the procedure for our purposes, we improved the original method and applied it to His-tagged proteins expressed as inclusion bodies in *E. coli*.

The 10 His-tagged proteins expressed as inclusion bodies were purified and solubilized with 8 M urea (see Materials and methods). The solubilized inclusion bodies were subsequently bound to Ni-NTA resin and subjected to refolding. Refolding was done on the column by changing buffers from denaturing 20 mM Tris–HCl, pH 8.0, buffer containing 8 M urea, to buffer (no urea) containing detergent and then buffer containing β-cyclodextrin. As a detergent in this particular refolding procedure, we used 0.1% Triton X-100, although different detergents can be tested. We are presently developing a mini-screen to identify which detergent(s) (non-ionic, ionic, and zwitterionic) might be best for a particular target protein (unpublished results). Adding 0.5 M NaCl to the detergent buffer allows removal of impurities nonspecifically bound to the resin. After the detergent was bound, the protein refolding was initiated by washing the column with β-cyclodextrin solution. The flow rate was kept at 0.5 ml/min. After washing the column with 20 mM Tris–HCl, pH 7.5, 100 mM NaCl to remove remaining β-cyclodextrin, the protein was eluted with buffer supplemented with 300 mM imidazole. The purity of target proteins in elution buffer was over 90%, as confirmed by SDS-PAGE (data not shown). After a polishing step (IEX or SEC), all refolded proteins were concentrated to 20 mg/ml or more without any occurrence of visible aggregation. We have been able to obtain 30–100% refolding in 7 out of 10 tested proteins. All refolded proteins were subjected to DLS analysis and five out of seven refolded proteins were monodisperse. Six of the seven refolded proteins were able to produce crystals of varying qualities. The yield of refolded proteins, DLS data and

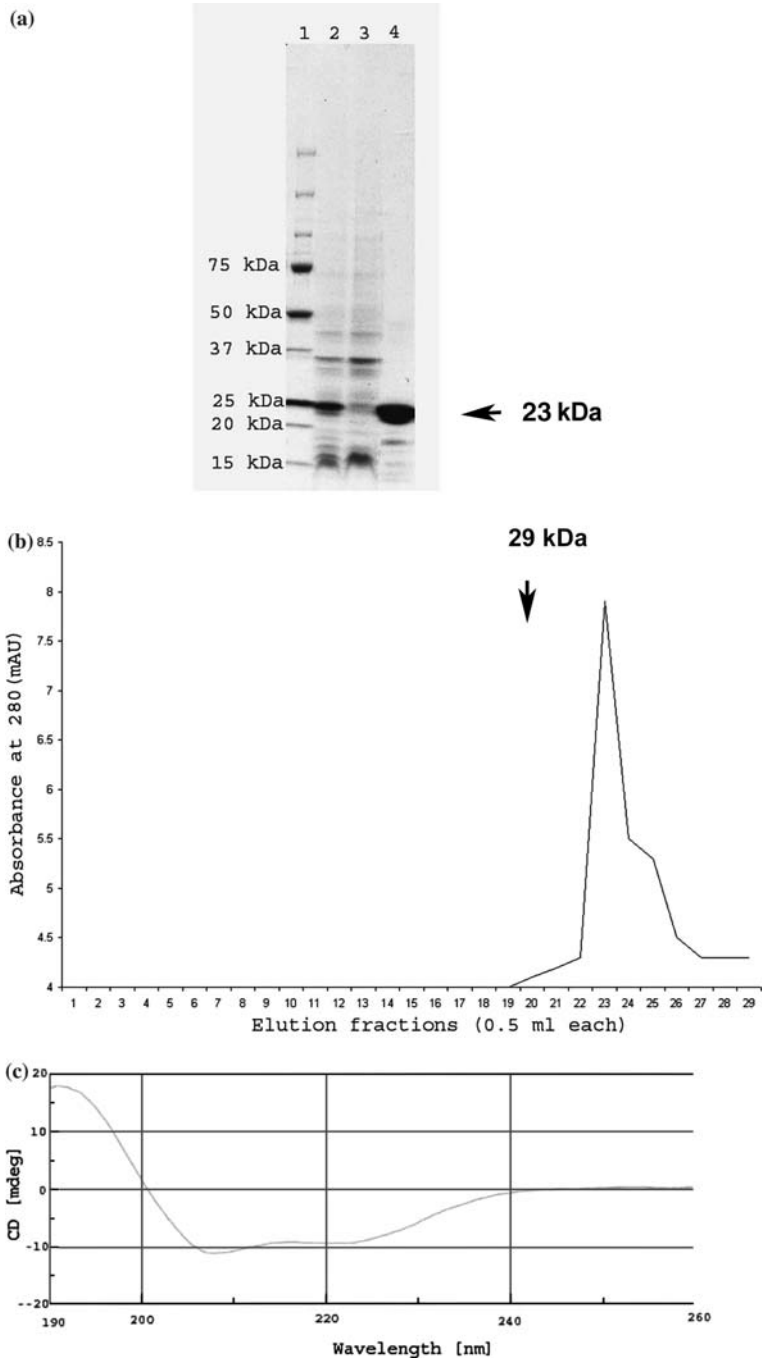

Figure 1. Analysis of BSGC target 1349B refolded and purified by on-column chaperone-like chemical refolding. (a) 4–20% SDS-PAGE analysis of refolded 1349B. Lane 1, protein molecular weight (MW) marker; lane 2, solubilized IB; lane 3, flow-through fraction from Ni-NTA column; lane 4, refolded 1349B eluted from Ni-NTA column (MW 23 kDa) indicated by arrow. (b) Fractionation of refolded 1349B by size exclusion chromatography. The protein eluted in a volume of 11.5 ml, indicating a relative molecular mass lower than 29 kDa. 1349B has a MW of 23 kDa, which is consistent with elution profile and confirms monomeric state of the refolded protein. (c) CD spectra of refolded 1349B. The protein concentration was 0.4 mg/ml in 10 mM Tris–HCl, pH 8.0, and 10 mM NaCl.

Table 1. Summary of the results obtained for BSGC refolded targets.

Targets	#cysteines	MW, kDa	% 'Refolded'[a]	DLS[b] (nm/% PD)	Crystallized
1049B	1	36	50	32.5/35.7%	Yes
1084B	1	40	40	Not measurable	Yes
1089B	2	17	100	2.48/26.1%	No
1105B	1	70	100	4/22.5%	3.2 Å data/solved
1113B	1	19	100	3.6/30.8%	Yes
1277B	4	44	100	4.46/10.5%	Yes
1294B	6	49	0	NA	–
1315B-2	4	61	0	NA	–
1338B-2	1	16	0	NA	–
1349B	0	23	80	3/20.7%	2.8 Å data

[a]% target eluted/% target loaded.
[b]nm, hydrodynamic radius in nanometers.
NA = not applicable.
PD = polydispersity.

success of crystallization efforts are presented in Table 1. In order to check for the presence of multimeric species formed during renaturation SEC can be used. For example, one of the BSGC targets, 1349B, expressed as an IB, was refolded with 80% success. Figure 1 shows analysis of this protein. According to SDS-PAGE, the eluted fractions contain >90% of target protein (Figure 1a). The DLS of refolded 1349B shows a hydrodynamic diameter of 3 nm and 20.7% of polydispersity (Table 1). To confirm its monomeric state, 0.4 mg of refolded protein was loaded on an analytical SEC (see Materials and methods). The elution profile showed the monodispersity of the analyzed protein (Figure 1b). CD data revealed well defined secondary structure (Figure 1c).

Figure 2 presents crystals obtained from two refolded targets – 1105B and 1349B. Data were collected from both crystals and the 1105B structure was solved (manuscript in preparation). From the crystal structure, we could not detect any detergent bound to the refolded protein, confirming that β-cyclodextrin efficiently strips the detergent from the protein during the refolding procedure.

Conclusion

Here we present a refolding method suitable for production of soluble proteins for structural studies. One of the key steps in structural genomics is the rapid production of purified native protein. The expression of recombinant proteins in transformed microorganisms is often hampered by the formation of insoluble protein

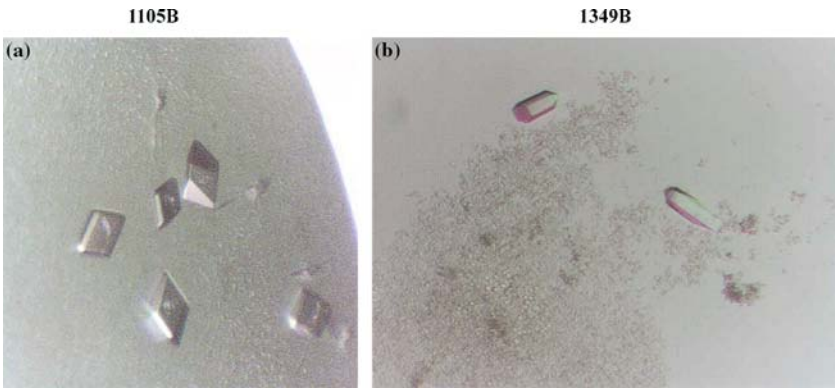

Figure 2. Crystals obtained from two BSGC targets refolded by on-column chaperone-like chemical refolding. (a) Crystals of 1105B, 300 × 150 × 150 μm. (b) Crystals of 1349B, 100 × 50 × 50 μm.

aggregates. Technology for refolding proteins that are expressed as inclusion bodies is still a major bottleneck in protein production. Such technology must be scale invariant, easily automated, applicable for a broad range of proteins, and economical. On-column chaperone-like chemical refolding meets these criteria. Column-based refolding with detergent and cyclodextrin promises to facilitate the rapid and efficient refolding of His-tagged recombinant proteins. The binding of denatured protein through a His-tag followed by a detergent wash efficiently prevents irreversible protein aggregation upon denaturant removal and significantly increases renaturation yield. It also eliminates the work of protein concentration, a significant time consuming step in the dilution-based approach. The proteins refolded by these methods produced crystals of various degrees of quality suitable for optimization and data collection. This can be easily automated and can become part of a soluble protein production pipeline for structure and functional studies.

Acknowledgements

We thank Hisao Yokota and Barbara Gold for cloning, Marlene Henriquez and Bruno Martinez for expression studies and cell paste preparation, Jarmila Jancarik and Ramona Pufan for crystallization of the proteins, Vaheh Oganesyan and Debanu Das for diffraction data collection, and John-Marc Chandonia for bioinformatics search of the genes. The work described here was supported by the National Institutes of Health GM 62412.

References

1. Yokoyama, S. (2003) *Curr. Opin. Chem. Biol.* **7**(1), 39–43.
2. Hou, J., Sims, G., Zhang, C. and Kim, S.-H. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 2386–2390.
3. Clark, E.B. (1998) *Curr. Opin. Biotechnol.* **9**, 157–163.
4. Yang, X.A., Dong, X.Y., Li, Y., Wang, Y.D. and Chen, W.F. (2004) *Protein Expr. Purif.* **33**, 332–338.
5. Shi, Y., Jiang, C., Chen, Q. and Tang, H. (2003) *Biochem. Biophys. Res. Comm.* **303**, 475–482.
6. Chang, S.-Y., Tsai, P.-C., Tseng, C.-S. and Liang, P.-H. (2001) *Protein Expr. Purif.* **23**, 432–439.
7. Rozema, D. and Gellman, S.H. (1996) *J. Biol. Chem.* **271**, 3478–3487.
8. Rozema, D. and Gellman, S.H. (1996) *Biochemistry* **35**, 15760–15771.
9. Daugherty, D.L., Rozema, D., Hanson, P.E. and Gellman, S.H. (1998) *J. Biol. Chem.* **273**, 33961–33971.
10. Nath, D. and Rao, M. (2001) *Eur. J. Biochem.* **268**, 5471–5478.
11. Dieckman, L., Gu, M., Stols, L., Donneley, M.I. and Collart, F.R. (2002) *Protein Expr. Purif.* **25**, 1–7.
12. Kim, R., Sandler, S.J., Goldman, S., Yokota, H., Clark, A.J. and Kim, S.-H. (1998) *Biotechnol. Lett.* **20**, 207–210.