The REFOLD database provides an annotated and analytical catalog of protein refolding methods

Michelle Chow, Ashley Buckle, James Whisstock and Stephen Bottomley*
Department of Biochemistry, Building 13D, Monash University, Clayton, Victoria 3800, Australia
* Corresponding author

Producing recombinant proteins as insoluble inclusion bodies is a relatively simple, efficient, and widely used technique to obtain large quantities of pure target protein. With inclusion bodies, the protein remains with the insoluble cell fraction, usually making up a high proportion of protein in that fraction, and is relatively easy to purify. The challenge comes in unfolding and properly renaturing each unique protein during purification. A myriad of published studies detail different refolding protocols, but until recently there has been no organized catalog of these methods. To meet this need, we developed the REFOLD database **(http://refold.med.monash.edu.au)**, a free and easily accessed resource that facilitates the ordered submission and search of purification and refolding protocols for recombinant proteins (1, 2).

REFOLD provides a repository for refolding protocols in a standardized and streamlined format. Each refolding protocol contains detailed information regarding expression conditions, buffers, methods, and conditions for solubilization and refolding, as well as structural and biochemical properties of the protein. With just over 200* records currently, and hopefully thousands in the future, REFOLD has the potential to become a predictive tool for scientists wishing to develop new refolding protocols for novel proteins.

Data entry in REFOLD

Access to and registration on REFOLD is free, and refolding data and protocols are easily entered online using a one-page form. The form is divided into three sections, which provide information on the protein, expression details, and the refolding method. The **protein** section contains information on the biochemical and structural properties of the protein, cross-references to the UniProt (3) and SCOP (4) databases, and a link to the reference in PubMed. The **expression** section includes information on the host organism and strain, expression vector, fusion tags, cell density at induction, and the time and temperature of expression. This section also has fields describing the method for harvesting inclusion bodies from the cell, the use of lytic agents, the solubility of the expressed protein, and any purification steps done prior to refolding the protein. The **refolding** section details the method used to refold the protein; the wash, solubilization, and refolding buffers used; and specific refolding conditions, such as pH, time, temperature, and the use of molecular chaperones. A dialog box provides space for a detailed protocol.

Searching and data analyses in REFOLD

Data in REFOLD can be searched using either a simple search on any term or an advanced search on more specific parameters. Search results provide a link for each protein to the full refolding report and protocol, and links for the various parameters (e.g., SCOP family, number of disulfides, solubility, refolding method) that lead to records for additional proteins and protocols sharing that attribute. REFOLD also

REFOLD Database web page search results for "phosphatase."

shows graphical analysis for different parameters, allowing the user to easily compare the most common methods and conditions used (e.g., refolding method, fusion construct, redox agent, pH, temperature). Links from these graphical analyses lead directly to records pertaining to that particular parameter.

Currently, some early trends can be observed. For instance, the most popular methods of refolding proteins is, first, by simple dilution into refolding buffer, followed by dialysis (Figure 1). More than a third of the proteins (~70 entries) are refolded in the presence of additives that assist refolding, such as arginine and glycerol, with only a few protocols using molecular chaperones. Also, when proteins are clustered according to their SCOP families, some protein groups are refolded within relatively narrow pH ranges (Figure 2) – specifically, the E-set domains of sugar-utilizing enzymes (pH 7.8-8.5), eukaryotic proteases (pH 8.5-8.8) and transforming growth factor-betas (pH 8.0-8.5).

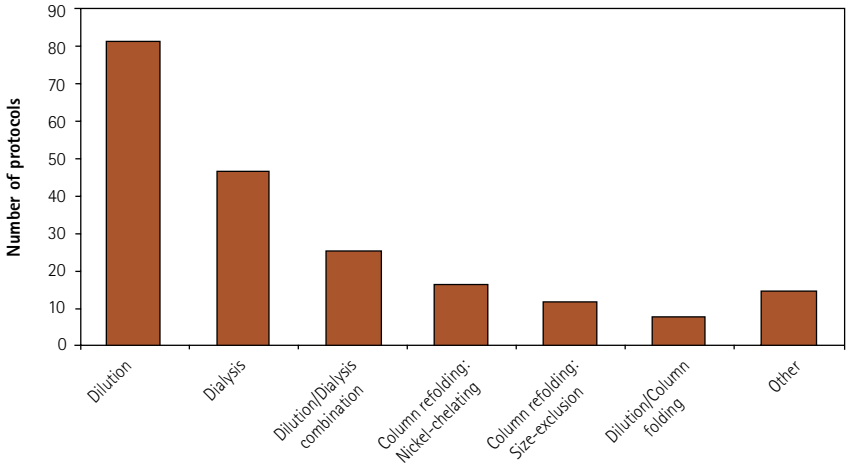

Figure 1. Frequency of refolding methods from data currently entered into REFOLD.

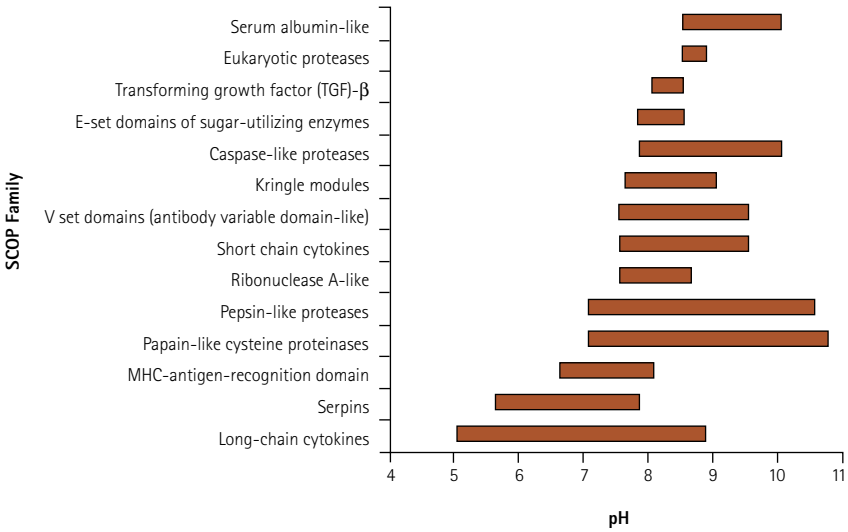

Figure 2. Refolding pH for proteins in REFOLD, clustered by SCOP family.

Value of REFOLD

REFOLD serves as a comprehensive catalog of refolding data presented in a logical, standardized format, allowing scientists easy access to published protocols and papers with extensive annotations. Each protein record contains a space for other users to add comments, providing a forum for the exchange of ideas and to generate discussion. Comments may pertain to the protocol, its application or adaptation to another protein, or the protein itself. The website also features a set of standard protocols for the most common refolding techniques. These generalized protocols provide a guide for novice protein refolders to design a protocol, which is then optimized for their particular protein.

Contributions to REFOLD

REFOLD can become a valuable tool for researchers, if protein scientists continue to contribute data. We welcome all published refolding protocols by scientists, who, in return, may benefit from the increased exposure of their published work. One of the most exciting aspects of REFOLD is its future potential to become a predictive tool in the design of new protocols, hopefully reducing the time required to establish protein purification protocols. We look forward to the continued growth and expansion of REFOLD through collaborative contributions from the scientific community. ■

REFERENCES

1. Buckle, A.M., Devlin, G.L., Jodun, R.A., Fulton, K.F., Faux, N., Whisstock J.C., and Bottomley, S.P. (2005) *Nat. Methods* **2**, 3.
2. Chow et al., (2005) *Protein Expr. Purif.*, In press
3. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., and Yeh, L.S. (2004) *Nucleic Acids Res.* **32** D115-119.
4. Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247** 536-540.

* Since this article was submitted, the number of records has increased to more than 300.