

## The matrix refolded

To the editor:

The postgenomic era has led to an explosion in protein expression and structural initiatives, from both a biotechnology and an academic perspective<sup>1,2</sup>. One of the most time- and cost-effective means of producing recombinant proteins uses genetically modified bacteria. A substantial hindrance to these efforts is the formation of insoluble inclusion bodies. In one study, only 13% of the human proteins tested could be purified from bacteria in soluble form<sup>3</sup>. To obtain native protein, it is necessary to purify inclusion bodies and subsequently solubilize them with denaturing agents. It is generally assumed that there is no one way to treat inclusion bodies and that a series of trial-and-error renaturation experiments must be performed. Identifying the optimal refolding conditions and methodology is therefore rate limiting. A wide range of protein refolding methodologies has been developed, using simple dilution as well as more complex matrix-assisted methods and the addition of solutes to renaturing buffers<sup>4–6</sup>.

The expansion of refolding methodologies is expected to continue in future years as structural biology strengthens and as the new field of structural genomics gains momentum. To exploit this wealth of

data, we have cataloged the methods used in the refolding of some 150 proteins in a web-accessible database, REFOLD. The efficiency of the outlined methods has been noted and annotation regarding protein structure has been included, such that the database can be searched via multiple parameters. As a data repository and experimental resource, the database allows new data to be rapidly deposited, validated and made available to the scientific community. Much valuable refolding data is never published—we aim to exploit this untapped resource by encouraging deposition in REFOLD. The database will thus become a powerful experimental tool for the optimization of protein renaturation, bypassing the slow and inefficient examination of the literature.

The database can be queried using numerous parameters, including gene species, refolding protocol and structural family (Fig. 1a). A spreadsheet-like list of results allows quick visualization of search results. The highly structured nature of REFOLD may facilitate the initial search for candidate refolding methodologies by allowing the inspection of protocols for proteins that have similar structural or functional properties.

Based on our current dataset, some trends evident in refolding protocols are worthy of mention. Refolding by simple dilution is by far the most common method reported in the literature, whereas more sophisticated matrix-assisted methods<sup>7</sup> and chaperone-assisted refolding<sup>8,9</sup> are less frequently reported (Fig. 1b). This highlights the importance of simplicity and cost in developing a viable renaturation strategy.

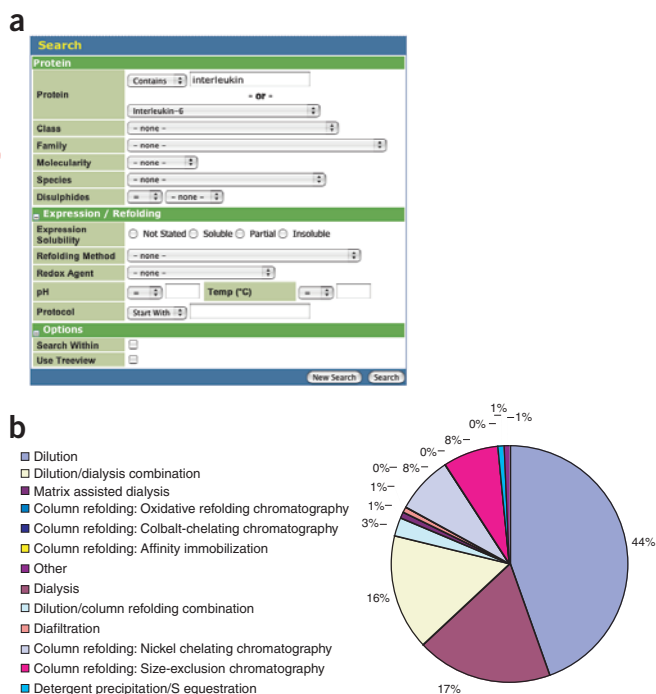
We hope that the utility of REFOLD will encourage both the deposition of refolding data in the same time frame as publication and the continual deposition of unpublished material. This will mean that the data become readily available to the community and amenable to analysis. In the long term, validation logic will be built into the deposition process, allowing a high degree of uniformity in the data structure, which will benefit experimentalists in data acquisition and handling.

REFOLD is freely available (<http://refold.med.monash.edu.au>).

Ashley M Buckle, Glyn L Devlin, Rachel A Jodun, Kate F Fulton, Noel Faux, James C Whisstock & Stephen P Bottomley

Department of Biochemistry and Molecular Biology, Faculty of Medicine and Victorian Bioinformatics Consortium, Monash University, Clayton Campus, Victoria 3800, Australia.

e-mail: [Ashley.Buckle@med.monash.edu.au](mailto:Ashley.Buckle@med.monash.edu.au) or [Steve.Bottomley@med.monash.edu.au](mailto:Steve.Bottomley@med.monash.edu.au)



**Figure 1** | The Refold Database (a) The web query interface to REFOLD. The database can be searched using multiple parameters relating to structural and chemical attributes. (b) Distribution of REFOLD entries according to refolding method.

1. Yee, A. *et al. Acc. Chem. Res.* **36**, 183–189 (2003).
2. Christendat, D. *et al. Prog. Biophys. Mol. Biol.* **73**, 339–345 (2000).
3. Braun, P. *et al. Proc. Natl. Acad. Sci. USA* **99**, 2654–2659 (2002).
4. Mayer, M. & Buchner, J. *Methods Mol. Med.* **94**, 239–254 (2004).
5. Middelberg, A.P. *Trends Biotechnol.* **20**, 437–443 (2002).
6. Clark, E.D.B. *Curr. Opin. Biotechnol.* **9**, 157–163 (1998).
7. Berdichevsky, Y. *et al. Protein Expr. Purif.* **17**, 249–259 (1999).
8. Altamirano, M.M., Golbik, R., Zahn, R., Buckle, A.M. & Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **94**, 3576–3578 (1997).
9. Gao, Y.G., Guan, Y.X., Yao, S.J. & Cho, M.G. *Biotechnol. Prog.* **19**, 915–920 (2003).