Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*

Tobias Cornvik[1,2], Sue-Li Dahlroth[1,2], Audur Magnusdottir[1,2], Maria Dolores Herman[1,2], Rosemarie Knaust[1], Monica Ekberg[2] & Pär Nordlund[2]

The implementation of efficient technologies for the production of recombinant mammalian proteins remains an outstanding challenge in many structural and functional genomics programs. We have developed a new method for rapid identification of soluble protein expression in *E. coli*, based on a separation of soluble protein from inclusion bodies by a filtration step at the colony level. The colony filtration (CoFi) blot is very well suited to screen libraries, and in the present work we used it to screen a deletion mutagenesis library.

The establishment of generic and efficient production systems of proteins suitable for structural and functional studies is one of the main challenges in current proteomics efforts as well as in commercial drug development and biotechnology programs[1,2]. Existing technologies for improving soluble protein expression in *E. coli* mainly depend on direct screening of expression parameters such as vectors, host strains and physical conditions[3,4]. Many mammalian proteins, however, cannot be expressed in a useful form in *E. coli* and often require the use of more expensive eukaryotic expression systems. Proteins expressed in eukaryotic hosts are harder to label, often have heterogenic covalent modifications and can be hard to produce in sufficient yields. *In vitro* evolution can be used to increase the probability of obtaining soluble protein expression in *E. coli*[5] but requires efficient methods to monitor protein solubility in many samples in parallel. Multiwell-based methods have recently been implemented for analysis of soluble expression in small liquid cultures, in which filtration, centrifugation and/or affinity purification are combined with dot-blot or SDS-PAGE analysis[4,6]. The general usefulness of these methods is limited by the relatively large costs and efforts of analyzing thousands of samples. Alternative methods of monitoring solubility in unlysed cells and on colony level have been described; these methods are based on a C-terminal fusion of a reporter protein, such as green fluorescent protein (GFP)[7] or chloramphenicol acetyltransferase[8]. An *in vivo* complementation assay, in which the relatively short N-terminal α peptide of β-galactosidase was used, has also been reported[9]. But fusion protein–based methods have some disadvantages; the fusion protein or protein fragment has to be removed before the protein of interest can be used for structural and biochemical studies. In addition, a fusion partner may affect the solubility of the target protein[5,10,11], and after removal the intrinsic solubility might be different than anticipated. Recently, a method claimed to have no effect on target solubility has been reported[12]. In this method, a split GFP is used to monitor protein solubility. A different approach, in which no fusion protein is used, has been described[13]. Soluble protein is separated from inclusion bodies by diffusion through an agarose layer, but only a limited number of colonies can be screened, and the general usefulness of the method still remains to be proven.

We have developed a method in which thousands of colonies can be analyzed. It does not rely on reporter protein fusions; instead a filtration step at the colony level is used, allowing efficient separation of soluble protein from inclusion bodies (**Fig. 1a** and **Supplementary Methods** online). The released proteins are captured on a nitrocellulose membrane and detected using standard immunochemicals. This makes the method universally applicable to any protein containing a suitable tag or epitope to which antibodies can be generated. The selected expression constructs only contain a small tag and can therefore, for most applications, go directly into scale-up.

The process is conceptually related to a 96 well–based method previously developed in our laboratory[6]. The method also turned out to have some similarities to previously developed filter sandwich methods used in directed evolution of antibodies[14] and enzymes[15]. The later methods, however, have never been shown to have any applicability in screening for soluble protein expression.

We validated the CoFi blot method by comparing it to centrifugation-based solubility analysis, the conventional method used to separate soluble proteins from inclusion bodies. We cloned a diverse set of 28 human and 4 mouse proteins (**Supplementary Table 1** online) as both N-terminal Flag- and His-tag fusions, in total 64 constructs. For detection and purification, a C-terminal His tag was present in all constructs. We dotted a colony containing each expression construct on an LB plate, after which the same colony was used to inoculate a liquid culture. The cells in liquid culture were grown, harvested, resuspended in lysis buffer and lysed by repeated freeze-thawing. We removed insoluble material by high-speed centrifugation and made dot blots of the cell lysate before and after centrifugation corresponding to total and soluble protein, respectively. We were able to detect total protein expression for 45 out of the 64 constructs.

[1]Department of Biochemistry and Biophysics, Stockholm University, S-106 91 Stockholm, Sweden. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden. Correspondence should be addressed to P.N. (par.nordlund@mbb.ki.se) or M.E. (monica.ekberg@mbb.ki.se).

PUBLISHED ONLINE 22 JUNE 2005; DOI:10.1038/NMETH767

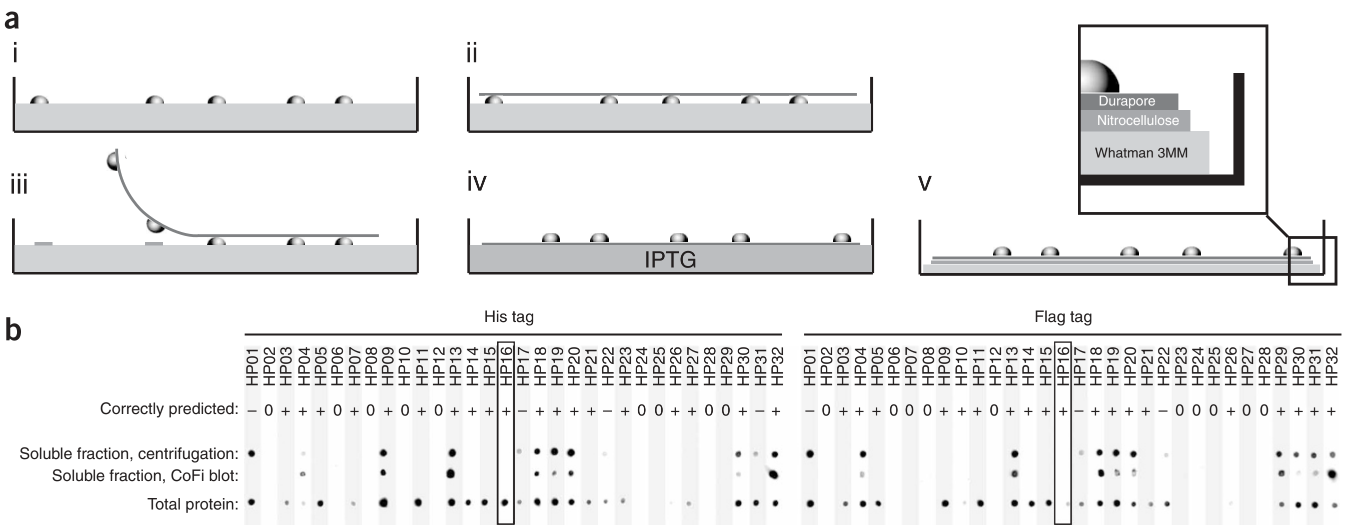

Figure 1 | Illustration and verification of the CoFi blot procedure (**a**) Schematic illustration of the CoFi blot procedure. Cells containing the expression constructs are plated and grown on an LB plate (i). The filter membrane is placed on top of the colonies (ii), which stick to the membrane and can be peeled off (iii). Protein expression is induced by placing the filter on an IPTG containing LB plate (iv). The cells are lysed, and soluble proteins diffuse through the filter membrane (Durapore) and are captured on the nitrocellulose membrane while inclusion bodies are retained on the filter membrane (v). By probing the nitrocellulose membrane with antibodies, colonies expressing soluble proteins can be identified. (**b**) CoFi blot results compared to dot blots after centrifugation. The CoFi blot results are in good correlation with the results from the centrifugation. We grew 32 different mammalian proteins in two different expression vectors in liquid culture and on plates. Colonies were CoFi blotted using the method described above. The cells grown in liquid culture were harvested by centrifugation, lysed using freeze-thawing, and insoluble material and inclusion bodies were removed by centrifugation. The remaining soluble fraction was dotted onto nitrocellulose membrane. As the CoFi blot is designed to discriminate between soluble protein and inclusion bodies, constructs with no total expression (no soluble protein or inclusion bodies) were excluded from the statistics, indicated by a zero.

In parallel, we subjected the colonies grown on plates to the CoFi blot method (**Fig. 1a**) and the blot had a remarkable similarity to the dot blot of soluble fractions after centrifugation (**Fig. 1b**). Using the CoFi blot, 38 out of 45 constructs (84%) were correctly predicted as soluble or insoluble, indicating that inclusion bodies can be separated from soluble protein using filtration at the colony level. Several constructs that were not predicted as soluble with the CoFi blot had the lowest levels of soluble protein expression in the centrifugation experiment (**Fig. 1b**). One of several possibilities is that the amount of soluble protein is below the detection limit of the CoFi blot.

To further verify the method, we tested 24 different *E. coli* proteins in both N-terminal Flag- and His-tag vectors as described above. Six constructs did not express total protein and were discarded. Of the remaining 42, 83% were correctly predicted as soluble or insoluble (**Supplementary Fig. 1** online).

In summary, we used 87 unique constructs that had detectable total protein expression to verify the CoFi blot method, and 84% of these constructs were correctly predicted as soluble or insoluble when compared to the data obtained using the centrifugation-based analysis. We have also shown that the method is reproducible by picking colonies of different intensities (strong, medium and weak) from a deletion library and reanalyzing them in quadruplicates with a new CoFi blot (**Supplementary Fig. 2** online). There is a good correlation between the first and second CoFi blot, and the average standard deviation in signal intensities between the quadruplicates is 4.5%.

To investigate the usefulness of the CoFi blot in screening *in vitro* evolution libraries, we created a deletion library (**Supplementary**

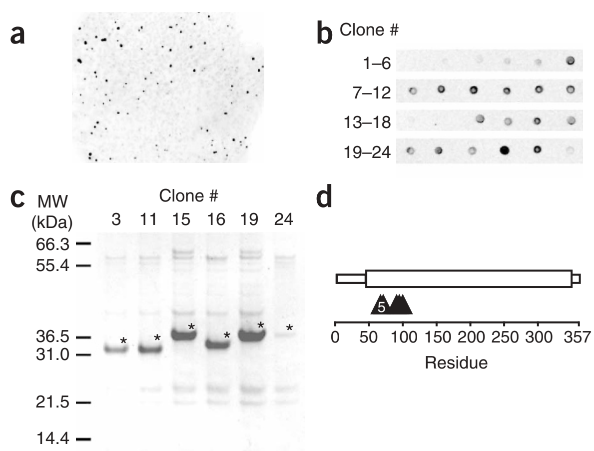

Figure 2 | CoFi blot screening of the deletion mutagenesis library of HP16. (**a**) A CoFi blot of the deletion library of HP16. Colonies expressing soluble protein are seen as black spots, which were located and picked for analysis. (**b**) Small-scale purification of 24 colonies identified as positive in the CoFi blot screen. Dot blots were made of the eluate and a large majority of the selected clones could be purified. (**c**) Six randomly selected clones were analyzed on Coomassie-stained 4–16% SDS-PAGE gels. * indicates that the band on the gel corresponds to the predicted size as determined from sequencing. (**d**) A schematic illustration of the predicted domain structure of HP16 according to PFAM. The large box represents the leucine carboxyl methyltransferase domain. Notably, when compared to the predicted domain, the translational start (▲) of the purifiable deletion mutants cluster close to the domain border. If there are multiple constructs that start with the same amino acid, the number of constructs is noted in the symbol.

Methods and **Supplementary Fig. 3** online) completely covering the open reading frame of HP16 (*LCMT1*; **Supplementary Table 1**), which did not produce a soluble protein as a full-length construct (**Fig. 1b**).

We screened the library for soluble expression using the CoFi blot (**Fig. 2a**) and picked 24 positive colonies, which were grown in liquid culture and subjected to small-scale Ni-NTA purification (**Fig. 2b** and **Supplementary Methods**). A large majority of the constructs could be purified and six random clones were analyzed on SDS-PAGE gels. The variations in the level of expression and the different lengths of the protein constructs are shown in **Figure 2c**. We also determined the translational start of nine of the constructs that we were able to purify, and notably, they cluster close to the start of the leucine carboxyl methyltransferase domain as predicted by PFAM[16], a database of domain alignments of protein families (**Fig. 2d**).

In conclusion, we have demonstrated that the CoFi blot method can be used to efficiently discriminate between colonies expressing soluble protein and those containing inclusion bodies. We have also shown that the CoFi blot can be used to screen *in vitro* evolution type libraries.

The use of the CoFi blot technology in the present work is only one of many potential applications of this new and efficient solubility screening technique. It is also well suited to screen other types of libraries intended for identifying soluble proteins, such as random mutagenesis libraries or libraries of expression vectors. Furthermore, the method is conceptually simple and remarkably robust.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank V. Lieu for help with purification and sequencing, and B. Engvall for cloning the targets for the benchmarking set. We also thank A. Beuscher, M. Högbom and P. Stenmark for their helpful comments on the manuscript. We would like to acknowledge the Swedish Research Council, the Wallenberg Consortium North, the Göran Gustafsson Foundation for Research in Natural Science and Medicine and the European Community–supported program Structural Proteomics in Europe (SPINE) for financial support.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 17 February; accepted 10 May 2005
Published online at http://www.nature.com/naturemethods/

1. Yokoyama, S. *Curr. Opin. Chem. Biol.* **7**, 39–43 (2003).
2. Edwards, A.M. *et al. Nat. Struct. Biol.* **7**, 970–972 (2000).
3. Hammarstrom, M., Hellgren, N., Van den Berg, S., Berglund, H. & Hard, T. *Protein Sci.* **11**, 313–321 (2002).
4. Braun, P. *et al. Proc. Natl. Acad. Sci. USA* **99**, 2654–2659 (2002).
5. Waldo, G.S. *Curr. Opin. Chem. Biol.* **7**, 33–38 (2003).
6. Knaust, R.K. & Nordlund, P. *Anal. Biochem.* **297**, 79–85 (2001).
7. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. *Nat. Biotechnol.* **17**, 691–695 (1999).
8. Maxwell, K.L., Mittermaier, A.K., Forman-Kay, J.D. & Davidson, A.R. *Protein Sci.* **8**, 1908–1911 (1999).
9. Wigley, W.C., Stidham, R.D., Smith, N.M., Hunt, J.F. & Thomas, P.J. *Nat. Biotechnol.* **19**, 131–136 (2001).
10. Sachdev, D. & Chirgwin, J.M. *J. Protein Chem.* **18**, 127–136 (1999).
11. Nomine, Y. *et al. Protein Expr. Purif.* **23**, 22–32 (2001).
12. Cabantous, S., Terwilliger, T.C. & Waldo, G.S. *Nat. Biotechnol.* **23**, 102–107 (2005).
13. Peabody, D.S. & Al-Bitar, L. *Nucleic Acids Res.* **29**, e113 (2001).
14. Dreher, M.L., Gherardi, E., Skerra, A. & Milstein, C. *J. Immunol. Methods* **139**, 197–205 (1991).
15. Heinis, C., Melkko, S., Demartis, S. & Neri, D. *Chem. Biol.* **9**, 383–390 (2002).
16. Bateman, A. *et al. Nucleic Acids Res.* **32**, D138–D141 (2004).