Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 346 (2005) 77–84
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Automated expression and solubility screening of His-tagged proteins in 96-well format

Renaud Vincentelli [a], Stéphane Canaan [a,b], Julien Offant [a], Christian Cambillau [a], Christophe Bignon [a,*]

[a] *Architecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS/Universités d'Aix-Marseille I et II, Case 932, 163 Avenue de Luminy, 13288 Marseille cedex 9, France*
[b] *Laboratoire d'Enzymologie Interfaciale et de Physiologie de la Lipolyse, CNRS-UPR 9025, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France*

Received 27 April 2005
Available online 15 August 2005

Abstract

A growing need for sensitive and high-throughput methods for screening the expression and solubility of recombinant proteins exists in structural genomics. Originally, the emergency solution was to use immediately available techniques such as manual lysis of expression cells followed by analysis of protein expression by gel electrophoresis. However, these handmade methods quickly proved to be unfit for the high-throughput demand of postgenomics, and it is now generally accepted that the long-term solution to this problem will be based on automation, on industrial standard-formatted experiments, and on downsizing samples and consumables. In agreement with this consensus, we have set up a fully automated method based on a dot–blot technology and using 96-well format consumables for assessing by immunodetection the amount of total and soluble recombinant histidine (His)-tagged proteins expressed in *Escherichia coli*. The method starts with the harvest of expression cells and ends with the display of solubility/expression results in milligrams of recombinant protein per liter of culture using a three-color code to assist analysis. The program autonomously processes 160 independent cultures at a time.

Keywords: Expression screening; Solubility screening; Structural genomics; Robotics; Dot–blot; Protein production; High-throughput

The past 5 years have witnessed a very fast evolution of technologies used in postgenomics. The driving force of this development has been the pressure undergone by techniques inherited from the pregenomics era to reach the throughput demand of postgenomics. In all cases, the throughput increase has resulted from the combined use of automation and of samples and consumables downsizing.

In structural genomics (SG),[1] the flow slows down at two stages: production of soluble proteins and crystallogenesis [1]. According to the above statement, we (and others) have streamlined initial stages of crystallogenesis, thanks to robotics and nano-drop technology [2,3]. In regard to soluble protein production, an international effort was undertaken during the same period, and numerous expression systems were improved [4]. This effort was so effective that it rapidly resulted in a dramatic increase in the number of samples to analyze with a predictable corollary: a growing need for

* Corresponding author. Fax: +33 4 91 26 67 20.
E-mail address: bignon@afmb.univ-mrs.fr (C. Bignon).

[1] *Abbreviations used:* SG, structural genomics; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; His, histidine; GFP, green fluorescent protein; HRP, horseradish peroxidase; TRX, thioredoxin; MBP, maltose binding protein; IPTG, isopropyl-β-D-thiogalactoside; PVDF, polyvinylidene fluoride; ECL, enhanced chemiluminescence; CCD, charge-coupled device; ORF, open reading frame; LB, Luria–Bertani.

0003-2697/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2005.07.039

sensitive and high-throughput methods for screening the expression and solubility of these numerous recombinant proteins.

Investigations in the setup of such screening methods can be roughly divided into two groups, depending on whether or not they make use of electrophoresis. The most widely used method remains sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) [5]. This technique offers undisputable advantages. It is easy to set up and provides valuable information on the apparent molecular mass of proteins. Eventually, its sensitivity can be increased by purifying recombinant histidine (His)-tagged proteins on Ni-affinity columns in a 96-well format prior to gel loading [6–10]. Automation of SDS–PAGE has also been tackled. For instance, gel loading/running can be performed by robots when horizontal gels are used (E-PAGE system, Invitrogen), and microfluidics electrophoresis (e.g., Agilent Technologies, Bartels Mikrotechnik) now allows fully automated processes to be envisaged.

Non-gel-based technologies have also been implemented. The green fluorescent protein (GFP) C terminal fused to target proteins has been evaluated as a folding reporter in vivo [11] and in vitro [12]. Another example of a soluble expression reporter system is provided by the FRETWorks S-Tag Assay (Novagen). Alternatively, some companies (e.g., Pierce, Sigma, Qiagen) propose to trap His-tagged proteins using Ni-coated plates in 96- and 384-well formats [13] and then to detect bound proteins by adding a Ni-activated peroxidase (INDIA HisProbe–HRP [horseradish peroxidase], Pierce) and a fluorescent or luminescent substrate. Finally, the protein mixture to analyze can be dot–blotted on a sheet of nitrocellulose. The recombinant protein is then detected by incubating the blot in the presence of a peroxidase-coupled antibody directed against a tag (generally a His6 tag) fused to the protein of interest. The light emitted by enzymatic degradation of a peroxidase luminescent substrate is recorded by autoradiography [14–16]. Because dot–blot processing resumes to pouring and decanting liquids, it should be expected to be a good candidate for automation. Until recently, however, this evolution was hampered by physical constraints such as vacuum manifold screwing/unscrewing and membrane handling. In contrast, steps located upstream of the dot–blot (cell lysis and recombinant protein purification) were successfully automated [16]. In the current study, we propose a solution to the above-mentioned limitations that results in a fully automated process for screening the expression and solubility of recombinant His-tagged proteins in *Escherichia coli*. The procedure starts with crude cell cultures and ends with the calculation and display of the amount of expressed protein in milligrams of total and soluble protein/culture volume.

Materials and methods

Protein expression

Coding sequences were inserted by recombination cloning (Gateway, Invitrogen) into different prokaryotic expression vectors, allowing N-terminal fusion with different tags [e.g., His6, thioredoxin (TRX), maltose binding protein (MBP), NusA]. In all cases, there was an N-terminal His tag. For expression, *E. coli* strains of the BL21(DE3) series (Novagen) were used. An overnight preculture was diluted 1:50 in 4 ml of fresh medium containing the required antibiotics in 24-well deep wells and grown at 37 °C. When the OD_{600} reached 0.6, protein expression was induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) at the final concentration of 0.5 mM and cells were grown at temperatures ranging from 25 to 42 °C until the OD_{600} reached a plateau (i.e., for an additional 3–4 h at 37 and 42 °C or overnight at 25 °C).

Cell recovery and lysis

Cells were recovered with magnetic beads. Two OD_{600} of cell suspension were loaded into a 96-well deep well. Buffers A and B (100 µl of each/well, Profos ColTrap Protein^{express}) were added and mixed by shaking. Trapped cells were concentrated for 5 min on a magnet (Qiagen), and culture medium was removed by pipetting. Away from the magnet, beads were resuspended in 200 µl of BugBuster HT (Novagen) and then incubated for 15 min at room temperature.

Cell lysate processing

Lysates were used either cleared (soluble fraction) or crude (total fraction).

Soluble fraction

Lysates (60 µl) were loaded into wells of a 0.65-µm pore size MADVNOB plate (Millipore), and the filtrate (soluble fraction) was recovered by vacuum aspiration.

Total fraction

Lysates (25 µl) were diluted to 150 µl with guanidinium buffer (50 mM Tris [pH 8], 150 mM NaCl, 10 mM imidazole, and 8 M guanidinium hydrochloride) and then filtered on 0.65 µm as above to remove any remaining particulate material. The filtrate (total fraction) was processed.

Dot–blot

Samples

After 60 µl (0.1 OD_{600}) of total fraction (40 samples) and 10 µl of the corresponding soluble fraction were

loaded into a 96-well microplate (sample plate), the remaining 16 wells were filled with proteins of the reference scale (see below). For the His6 tag to be freely accessible to antibodies, proteins were denatured by adjusting the volume to 200 µl with guanidinium buffer.

Reference scale

A 16-point calibration scale obtained by diluting a known amount of His-tagged protein in guanidinium buffer was loaded into the sample plate. This protein was either homemade or commercial (6×His protein ladder, Cat. No. 34705, Qiagen).

Dot–blot

Liquids were added by manual or automatic pipetting and were removed by aspiration through a MultiScreen vacuum manifold (MAVM 096 OR, Millipore) or a robot-driven T-Vac vacuum system (Tecan). Polyvinylidene fluoride (PVDF) filters of 0.45-µm pore size 96-well MultiScreen MSIPNOB plates (Millipore) were activated for 5 min with 100 µl of 70% ethanol, followed by 100 µl of water washout. The 96 samples were transferred from the sample plate to MSIPNOB plates and incubated for 30 min. After aspiration, membranes were washed one time with 200 µl of 10 mM Tris (pH 7.5), 150 mM NaCl, and 0.1% Tween 20 (TBST). Unoccupied binding sites were saturated for 30 min by incubating in the presence of 100 µl/well of 0.5% (w/v) blocking reagent in blocking buffer (Qiagen) and 0.1% Tween 20. Membranes were washed once with 200 µl of TBST and then incubated in the presence of 50 µl of anti-penta His–HRP conjugate antibody (Qiagen) diluted 1:2000 in blocking buffer. Then, 30 min later, membranes were washed four times with 200 µl of TBST and 50 µl of extemporaneously prepared ECL+ reagent (Amersham Biosciences) was added to each well. After 3 min of incubation, the enhanced chemiluminescence (ECL) solution was aspirated and the light emission was recorded for 1 to 5 min by a charge-coupled device (CCD) camera (Kodak LumiImager) or for 1 s/well by a robot-driven microplate reader (Tecan GENios Plus). The whole procedure took approximately 130 min.

Automation

The program processes 160 independent cultures at a time. In practice, the robot autonomously processes cultures 1–80 and then cultures 81–160. The processing of 80 cultures is described below.

Robot

The setup consisted of a Tecan Genesis Freedom 200 robot with an eight-needle pipetting arm, a handling arm for moving microplates, a Tecan GENios Plus microplate reader, a magnet for 96-well plates, and shaking and vacuum units [1].

Cell processing

Four 24-well deep wells (80 cultures) containing 4 ml/well of cell culture were positioned on the robot working bench. The machine transferred 100 µl of culture medium and then a 100-µl aliquot of cell suspension from each well into a clear 96-well microplate. The latter was moved by the robot arm to the microplate reader located below, and the OD_{600} of each well was measured. On the basis of this OD_{600}, the robot transferred two OD_{600} from the four 24-well deep wells to a single 96-well deep well. For OD_{600} less than 1.2 and greater than 20, the robot took 1800 and 100 µl, respectively. Cell recovery and lysis were performed as above. Of 200 µl of lysate in the first 40 samples, 30 µl was mixed by the robot with 10 µl of 4×SDS–PAGE sample buffer and the remaining volume was used to prepare total and soluble fractions as described above, except that 40 total and soluble fractions were directly recovered in a microplate by vacuum aspiration. Then 30 µl of each soluble fraction was mixed with 10 µl of 4×SDS–PAGE sample buffer. In another plate, samples and reference scale were prepared as described above. The same program was repeated for cultures 41–80.

Dot–blot

The contents of two sample plates were transferred by the robot to two ethanol-activated 0.45-µm pore size 96-well MSIPNOB microplates, which were processed as described above. After ECL removal, the machine sequentially transferred the two plates to the microplate reader and the amount of light emitted by each dot was recorded. Protein expression was assessed by a four-step calculation. First, the background [light signal produced by the lowest protein load of the reference scale (0 ng)] was deduced from all raw values. Second, after background removal, the data were translated in milligrams of protein by linear regression using the reference scale. Third, because the latter was made of pure His-tagged protein, the linear regression underestimated the expression; therefore, this first estimate was multiplied by a correction factor to take into account the influence of cell lysate (see Results and discussion). Fourth, the amount of protein was expressed as milligrams of protein per liter of culture by dividing the amount of protein by 0.1 (0.1 OD_{600} was the amount of bacteria effectively used for the dot–blot out of the two OD_{600} that had been processed) and multiplying by the total number of OD_{600}/L of culture determined on the basis of the OD_{600} recorded at the beginning of the experiment. Finally, to help in data analysis, the results were automatically displayed using a color code indicating the level of expression and solubility of each target. At this stage, the apparent molecular weight of the protein was eventually checked by SDS–PAGE using the samples saved by the robot for this purpose.

A synopsis of the complete protocol is given in Fig. 1.

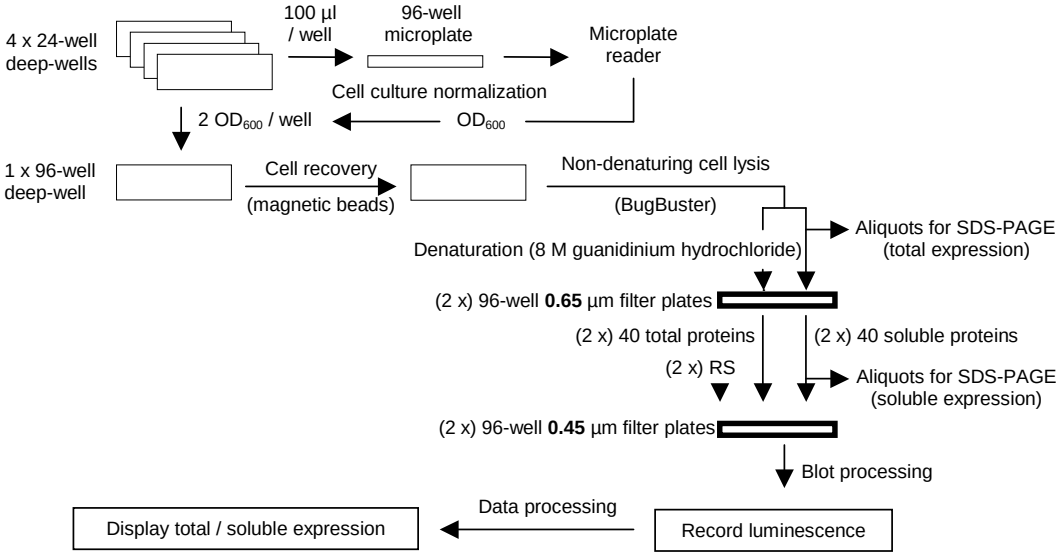


Fig. 1. Synopsis of the automated process for 80 cultures. RS, reference scale. For details, see Materials and methods.

Results and discussion

Handmade dot–blot using a CCD camera

The launch of disposable 96-well plates with a porous bottom made of 0.45-µm pore size nitrocellulose or PVDF (e.g., Millipore, Whatman, Pall), in which different reagents could be incubated and then removed from beneath by online vacuum, made it possible to combine the advantages of a dot–blot manifold and of automation.

To be used in a protein assay, these plates should fulfill the following criteria. First, the ECL signal should be at least as good as that obtained with a sheet of nitrocellulose sandwiched in a standard vacuum manifold. In particular, the light signal should be even throughout the entire surface of the plate. Second, the sensitivity of the assay should be at least that of SDS–PAGE.

The first point in the preceding paragraph was assessed by loading sixfold in a single plate a serial dilution of a known amount of pure His-tagged protein and then processing the plate (Fig. 2A). To accurately evaluate the reproducibility of the light signal emitted by each dilution, the mean value and standard deviation of the light signal of each dilution was calculated and plotted against the amount of loaded protein (Fig. 2B). Two types of information could be inferred from Fig. 2B. First, the extent of standard deviation, which accounts for the reproducibility of ECL signal produced by the same amount of protein loaded in six different wells of the plate, was low enough to be considered compatible with the use of these plates in a protein expression assay, and this was true for the entire range of the scale (0–2000 ng). Second, the point distribution followed a nonlinear regression that denoted saturating signals at high protein load. These were not accounted for by a saturation of the CCD camera given that a 1-min exposure revealed the same curve, although the plateau was in a range ($\sim 5 \times 10^5$) where no saturation appeared in the 5-min exposure curve. The plateau, therefore, indicated a membrane overloading. In contrast, when the reference scale was limited to 500 ng instead of 2000 ng, there was a linear relationship between the light signal and the protein load (Fig. 2C) that could be used in the linear regression analysis of a protein assay.

The second point above was addressed by analyzing the same protein samples by dot–blot and SDS–PAGE. A single open reading frame (ORF, Rv1096) was expressed under different experimental conditions defined by a fractional factorial approach [15]. The parameters combined in this approach included four N-terminal fusions (His6 tag, MBP, TRX, and NusA), four *E. coli* strains, three culture media, and two temperatures (Fig. 2D, III). At the end of the experiment, total and soluble expressions were analyzed by SDS–PAGE and dot–blot (Fig. 2D, I and II). Rv1096 is a *Mycobacterium tuberculosis* coding sequence, an organism whose genes are known to be difficult to express as soluble proteins in *E. coli* [1]. The use of MBP, TRX, and NusA was an attempt to address this issue [17]. As expected, SDS–PAGE analysis revealed that MBP and NusA fusions permitted a high level of soluble expression at low temperature regardless of the *E. coli* strain or growth medium used. In contrast, His6 and TRX were unable to enhance the solubility of Rv1096. Regarding correlation with dot–blot, a rapid survey indicated a good overall fit of the results provided by the two techniques. However, a closer examination of the results obtained with the two smallest fusions His6 tag and TRX (Figs. 2D, I) showed that the dot–blot revealed low levels of soluble expression that SDS–PAGE did not (A8, B8, D8, and F8). This signal was specific given that it was not

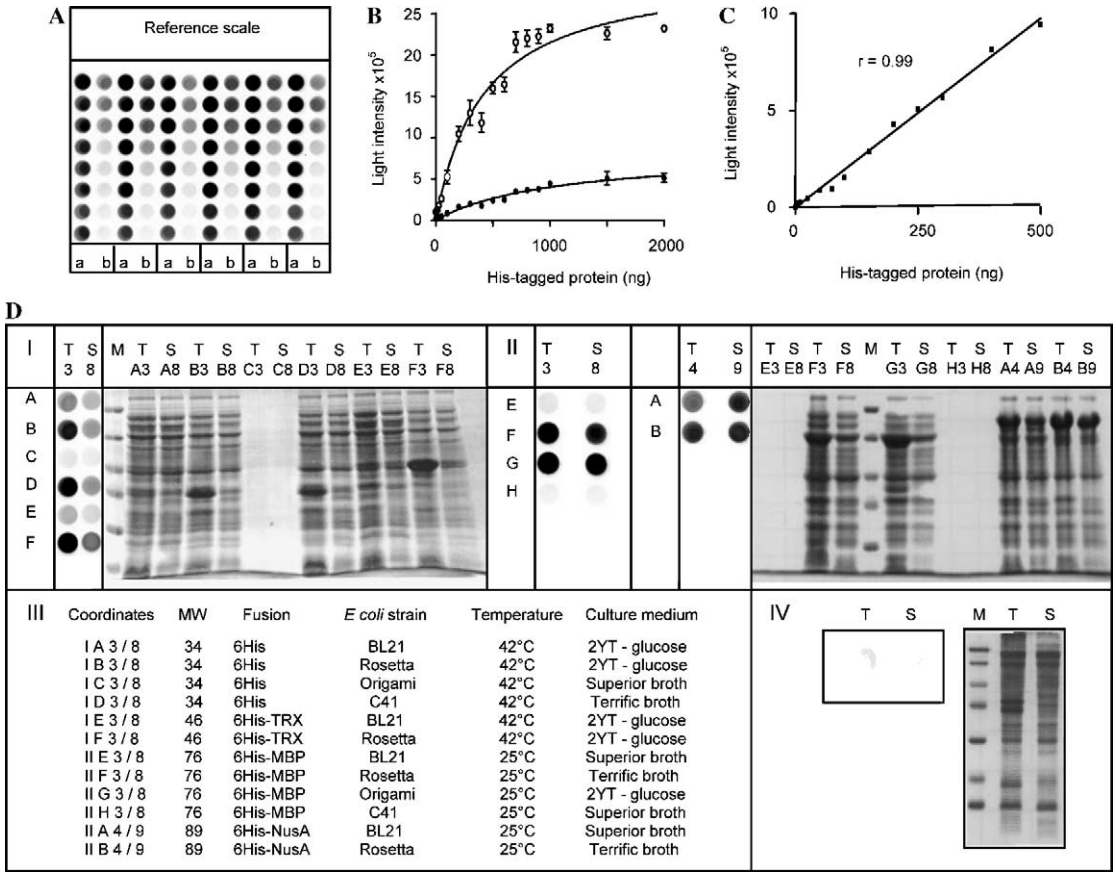


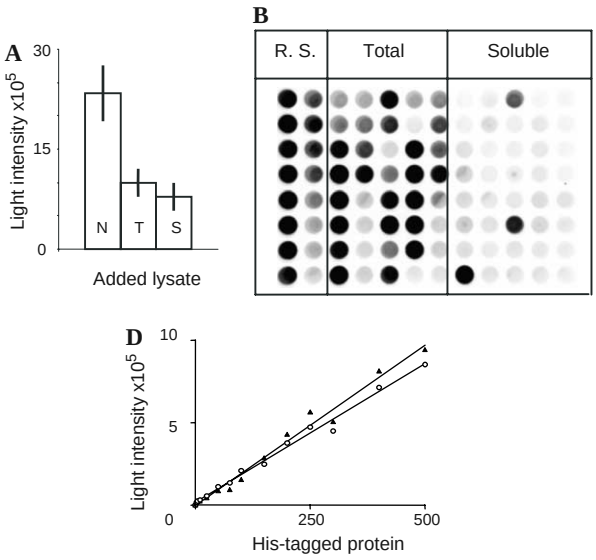
Fig. 2. Dot–blot calibration. (A) Dilutions of the His-tagged protein of the reference scale were loaded six times on the whole surface of a 96-well 0.45-μm filter plate. The plate was processed manually using a specific vacuum device (Millipore), and the ECL signal was recorded by a CCD camera. The picture displays the results obtained after a 5-min exposure time. Column a (top to bottom): 2000, 1500, 1000, 900, 800, 700, 600, and 500 ng; column b (top to bottom): 400, 300, 200, 100, 50, 25, 12.5, and 0 ng. (B) The mean value and standard deviation of the light signal of each dilution displayed in (A) was calculated and plotted against the amount of loaded protein. The points were linked following a nonlinear regression. The results of 1-min (black circles) and 5-min (open circles) recordings are shown. Light intensities are in arbitrary units. (C) This panel is the same as (B) (5-min recording time) except that the protein load ranged from 0 to 500 ng. The resulting points were linked following a linear regression. The confidence factor is indicated. (D) This panel gives a comparison of the results provided by dot–blot and SDS–PAGE. Total (T) and soluble (S) *E. coli* expression of the *M. tuberculosis* Rv1096 gene was evaluated under different experimental conditions defined by a factorial approach. At the end of the experiment, cells were lysed and the presence of the protein of interest in the total lysate or in the soluble fraction was checked by SDS–PAGE and dot–blot. I and II: to allow a direct comparison, whole gels are displayed, but only those parts of the dot–blot corresponding to the samples that have been run on gels are shown. Each dot is defined by its position in the 96-well plate, a combination of columns 3, 4, 8, and 9 and rows A to H. This combination is recalled above the corresponding gel track (e.g., A3, A8). M: molecular weight markers (116, 66, 45, 35, 25, 18, and 14 kDa). I: His and His–TRX fusions. II: His–MBP and His–NusA fusions. III: reporting of various experimental conditions of the factorial map, with corresponding dot–blot/gel coordinates, and expected molecular weights (in kDa) of each fusion, given in the first and second columns, respectively. IV: tuner(DE3)pLysS cells transformed with an expression plasmid bearing a target gene that did not express in *E. coli* processed as above.

detected when no protein was present because the culture did not grow (Figs. 2D, I, C3/8; II, E3/8 and H3/8) or because no protein was expressed (Fig. 2D, IV, T and S). This higher sensitivity of the dot–blot is of particular importance when proteins are to be used in structural studies because small tags such as His6 generally do not need to be removed prior to crystallogenesis, whereas this is not the case for large fusions such as MBP and NusA. Finally, when the protein of interest happens to comigrate with endogenous *E. coli* proteins, detection by SDS–PAGE requires larger overexpression than dot–blot to distinguish it from the background. This was the case for MBP fusion of Rv1096 (Fig. 2D, II, F3/8 and G3/8).

In conclusion, 0.45-μm filter plates satisfied both the reproducibility and sensitivity criteria defined above and, therefore, could be integrated in an automated screening procedure.

Automated quantitative dot–blot using a microplate reader

A fully automated process required two additional implementations.



C

Project: TB 37°C Plate: 1									
Date: 28/11/03									
User: RV									
Total expression (mg/l culture)					Soluble expression (mg/l culture)				
2.8	2	21	2.2	4.4	-	-	8.5	-	-
4.8	6.2	13.2	-	9.4	0.0	0.0	0.0	-	-
26	9	0.2	32	7	0.0	-	0.0	0.0	-
57	18.9	5	34.7	15.8	0.5	-	0.0	0.0	-
38	1.5	19.1	33.4	4.1	0.2	-	0.0	0.2	0.0
27.5	0.6	43.1	34.5	0.8	1.1	0.2	13.6	0.4	0.0
20.6	0.2	5.8	34.3	1.1	0.0	0.0	-	0.0	0.0
39.1	0.8	16.2	0.0	0.2	35.6	0.0	-	-	0.0

Expression and solubility				
1	9	17	25	33
2	10	18	26	34
3	11	19	27	35
4	12	20	28	36
5	13	21	29	37
6	14	22	30	38
7	15	23	31	39
8	16	24	32	40

	≤ 0.5 mg/l
	0.5 < x < 2 mg/l
	≥ 2mg/l
	No expression
	Not soluble
	Soluble ≤ 2mg/L
	Soluble > 2mg/L

Fig. 3. Automated and quantitative dot–blot. (A) This panel shows the effect of cell lysate on ECL signal. The three histograms represent the mean value of the light intensity (in arbitrary units) produced by processing in a dot–blot experiment 200 ng of His-tagged protein of the reference scale in the presence of 0 µl (N, 32 replicates) or 10 µl of lysate (0.1 OD_{600}) of nontransformed *E. coli* cells (T, 24 replicates) or its soluble part (S, 24 replicates). Standard deviations are indicated by vertical bars at the top of the histograms. (B) A total of 40 *E. coli* cultures expressing different recombinant proteins were lysed, and 0.1 OD_{600} of the total or soluble fraction of each lysate was dot–blotted. At the end of the experiment, light signals were recorded by the CCD camera. R.S., reference scale; Total, total expression; Soluble, soluble expression. (C) After reading with the CCD camera, the plate was scanned using the Tecan microplate reader. Values below the background are indicated by a dash (–); these did not exceed −0.1. (D) Light signals emitted by the reference scale and recorded by the CCD camera (open circles) and by the microplate reader (open triangles) were plotted against the amount of loaded protein.

Influence of cell lysate

Measuring the His-tagged protein content of an *E. coli* lysate by directly comparing the ECL signal produced by a well containing an aliquot of this lysate with that produced by a well bearing a pure His-tagged protein would underestimate the former [18]. Thus, for the comparison to be quantitative, a correction factor was required.

In the following experiment, an amount of His-tagged protein of the reference scale chosen in the middle of the scale (200 ng) was mixed with 0 or 10 µl of total lysate of nontransformed *E. coli* cells or its soluble part. Each of these three experimental conditions was loaded as several replicates in wells of a single plate (32 wells for the undiluted protein, 24 wells for the protein diluted with the total lysate, and 24 wells for the protein diluted with the soluble fraction of the lysate) and dot–blotted. The light signal was recorded, and the mean value and standard deviation were calculated in each of the three cases. As shown in Fig. 3A, total and soluble lysates exhibited comparable binding displacement capacity, although the soluble fraction reproducibly appeared to be slightly more potent than the total fraction for an unknown reason. The correction factor, defined as the ratio of signal intensities measured in the absence or presence of lysate, was estimated to be 2.0 and 2.4 for total lysate and soluble lysate, respectively. Taking into account the amplitude of standard deviation of the experiment defining the correction factor, an averaged 2.2 correction factor for both total and soluble lysates was considered suitable for use in our assay.

Automation of luminescent signal recording and treatment

Integrating a CCD camera on the robot was impractical, so we decided to use the microplate reader located below the robot bench and driven by the robot computer.

In the experiment reported in Figs. 3B–D, ECL signals recorded by the remote CCD camera used in the experiments reported in Fig. 2 and by the robot's microplate reader were compared. To that end, 40 *E. coli* cultures expressing different His-tagged *M. tuberculosis* proteins were processed as usual. After ECL removal, the plate was scanned using the CCD camera and then the microplate reader. The raw data of the latter were further processed, and the results were displayed using a color code as described in Materials and methods. The raw data obtained with the CCD camera and the processed data obtained with the microplate reader are reported in Figs. 3B and C, respectively. Note that the reference scale visible in Fig. 3B did not appear in the layout of the results provided by the microplate reader because it has been used for calculating the protein concentrations displayed in the boxes representing each well of the microplate and, therefore, could be omitted. The lysate correction factor was used to compensate for the reference scale underrating, as discussed earlier.

A comparison of the results reported in Figs. 3B and C indicated a good fit between the data obtained with the CCD camera and those obtained with the microplate reader. This correlation was further confirmed by plotting the light intensity emitted by the reference scale and recorded by the two devices against the amount of loaded protein. The same slope was obtained, indicating that a microplate reader perfectly substituted for a CCD camera in a quantification purpose (Fig. 3D).

These results were typical of those obtained when expressing *M. tuberculosis* ORF in *E. coli* under basic conditions (no N-terminal fusion to enhance solubility and expression in Luria–Bertani (LB) medium at 37 °C). Only three proteins (8, 17, and 22 in the "expression and solubility" panel of Fig. 3C, corresponding to the three main spots in the "soluble" panel of Fig. 3B) were expressed soluble at levels compatible with a direct scale-up without further improving expression conditions ($\geqslant$2 mg/L culture). A single target (6) was expressed soluble at less than 2 mg/L, requiring either large culture volumes or additional screening for better expression conditions. Only five targets (15, 19, 26, 32, and 40) were not expressed either soluble or insoluble and, therefore, required trying other expression conditions. Finally, most targets (1–5, 7, 9–14, 16, 18, 20, 21, 23–25, 27–31, and 33–39) were expressed insoluble at levels compatible with inclusion bodies refolding screening [19].

The idea of using a color code came from the observation that raw data provided by a CCD camera (gray scale) or by a microplate reader (unitless numbers) depended on recording times and, hence, could be misleading in case of saturating or weak signals. We reasoned that this was less likely to occur if the decision of whether or not to proceed with a given target relied on a color code based on absolute information (e.g., mg/ml) and not on relative information (e.g., gray scale, unitless numbers). In addition, whereas Figs. 3B and C provide the same information, the access time to this information is shorter in the latter than in the former.

The reliability limit of dot–blots in expression screening comes from the detection method (immunological recognition of a His6 tag). The assay is unable to distinguish between full-length and truncated proteins so long as the His tag is not degraded. The solution is to use techniques providing information on the actual protein size such as capillary electrophoresis, SDS–PAGE, and Western blotting. However, even in this case, we believe that it is worth using a dot–blot screening first because (i) we found this event to be very marginal in the course of four SG programs (unpublished data); (ii) only samples tested positive by the machine will be analyzed on gel, and this may significantly reduce the amount of work; and (iii) aliquots are saved by the robot for this purpose. Alternatively, the use of a C-terminal His tag (pET-DEST42 expression vector) will preclude detecting proteins whose truncation is due to premature translation termination.

Acknowledgments

We thank Denis Chaix for his contribution to the initial steps of this work and Arie Geerlof (EMBL) for the gift of the Gateway expression plasmids encoding thioredoxin and NusA fusions. This work was funded by the SPINE, VIZIER, and X-TB European programs as well as by Marseille–Nice Genopole.

References

[1] R. Vincentelli, C. Bignon, A. Gruez, S. Canaan, G. Sulzenbacher, M. Tegoni, V. Campanacci, C. Cambillau, Medium-scale structural genomics: strategies for protein expression and crystallization, Acc. Chem. Res. 36 (2003) 165–172.

[2] G. Sulzenbacher, A. Gruez, V. Roig-Zamboni, S. Spinelli, C. Valencia, F. Pagot, R. Vincentelli, C. Bignon, A. Salomoni, S. Grisel, D. Maurin, C. Huyghe, K. Johansson, A. Grassick, A. Roussel, Y. Bourne, S. Perrier, L. Miallau, P. Cantau, E. Blanc, M. Genevois, A. Grossi, A. Zenatti, V. Campanacci, C. Cambillau, A medium-throughput crystallization approach, Acta Crystallogr. D Biol. Crystallogr. 58 (2002) 2109–2115.

[3] D. Hosfield, J. Palan, M. Hilgers, D. Scheibe, D.E. McRee, R.C. Stevens, A fully integrated protein crystallization platform for small-molecule drug discovery, J. Struct. Biol. 142 (2003) 207–217.

[4] S. Yokoyama, Protein expression systems for structural genomics and proteomics, Curr. Opin. Chem. Biol. 7 (2003) 39–43.

[5] Y.-P. Shih, W.-M. Kung, J.-C. Chen, C.-H. Yeh, A.H.-J. Wang, T.-F. Wang, High-throughput screening of soluble recombinant proteins, Protein Sci. 11 (2002) 1714–1719.

[6] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.-A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, Proc. Natl. Acad. Sci. USA 99 (2002) 11664–11669.

[7] K. Büssow, C. Quedenau, V. Sievert, J. Tischer, C. Scheich, H. Seitz, B. Hieke, F.H. Niesen, F. Götz, U. Harttig, H. Lehrach, A catalog of human cDNA expression clones and its application to structural genomics, Genome Biol. 5 (2004) R71.

[8] M. DiDonato, A.M. Deacon, H.E. Klock, D. McMullan, S.A. Lesley, A scaleable and integrated crystallization pipeline applied to mining the *Thermotoga maritima* proteome, J. Struct. Funct. Genomics 5 (2004) 133–146.

[9] G.E. Folkers, B.N.M. van Buuren, R. Kaptein, Expression screening, protein purification, and NMR analysis of human protein domains for structural genomics, J. Struct. Funct. Genomics 5 (2004) 119–131.

[10] C. Scheich, V. Sievert, K. Büssow, An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography, BMC Biotechnol. 3 (2003) 12.

[11] K. Omoya, Z. Kato, E. Matsukuma, A. Li, K. Hashimoto, Y. Yamamoto, H. Ohnishi, N. Kondo, Systematic optimization of active protein expression using GFP as a folding reporter, Protein Expr. Purif. 36 (2004) 327–332.

[12] M.A. Coleman, V.H. Lao, B.W. Segelke, P.T. Beernink, High-throughput, fluorescence-based screening for soluble protein expression, J. Proteome Res. 3 (2004) 1024–1032.
[13] K. Fineberg, T. Fineberg, A. Graessmann, N.W. Luedtke, Y. Tor, R. Lixin, D.A. Jans, A. Loyter, Inhibition of nuclear import mediated by the Rev-arginine rich motif by RNA molecules, Biochemistry 42 (2003) 2625–2633.
[14] R.K. Knaust, P. Nordlund, Screening for soluble expression of recombinant proteins in a 96-well format, Anal. Biochem. 297 (2001) 79–85.
[15] C. Abergel, B. Coutard, D. Byrne, S. Chenivesse, J.-B. Claude, C. Deregnaucourt, T. Fricaux, C. Gianesini-Boutreux, S. Jeudy, R. Lebrun, C. Maza, C. Notredame, O. Poirot, K. Suhre, M. Varagnol, J.M. Claverie, Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets, J. Struct. Funct. Genomics 4 (2003) 141–157.
[16] H. Nguyen, B. Martinez, N. Oganesyan, R. Kim, An automated small-scale protein expression and purification screening provides beneficial information for protein production, J. Struct. Funct. Genomics 5 (2004) 23–27.
[17] M. Hammarström, N. Hellgren, S. Van den Berg, H. Berglund, T. Härd, Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*, Protein Sci. 11 (2002) 313–321.
[18] V. Kery, J.R. Savage, K. Widjaja, B.K. Blake, D.R. Conklin, Y.S. Ho, X. Long, M. von Rechenberg, T.I. Zarembinski, J.J. Boniface, Expression screen by enzyme-linked immunofiltration assay designed for high-throughput purification of affinity-tagged proteins, Anal. Biochem. 317 (2003) 255–258.
[19] R. Vincentelli, S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, C. Bignon, High-throughput automated refolding screening of inclusion bodies, Protein Sci. 13 (2004) 2782–2792.