



Dual Expression System Suitable for High-Throughput Fluorescence-Based Screening and Production of Soluble Proteins

Sandrine Braud, Mireille Moutiez, Pascal Belin, Nicolas Abello, Pascal Drevet, Sophie Zinn-Justin, Marie Courçon, Cédric Masson, Janie Dassa, Jean-Baptiste Charbonnier, et al.

► To cite this version:

Sandrine Braud, Mireille Moutiez, Pascal Belin, Nicolas Abello, Pascal Drevet, et al.. Dual Expression System Suitable for High-Throughput Fluorescence-Based Screening and Production of Soluble Proteins. *Journal of Proteome Research*, 2005, 4 (6), pp.2137-2147. <10.1021/pr050230i>. <hal-04135839>

HAL Id: hal-04135839

<https://hal.science/hal-04135839v1>

Submitted on 28 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Dual Expression System Suitable for High-Throughput Fluorescence-Based Screening and Production of Soluble Proteins

Sandrine Braud,[†] Mireille Moutiez,[†] Pascal Belin, Nicolas Abello,[‡] Pascal Drevet, Sophie Zinn-Justin, Marie Courçon, Cédric Masson, Janie Dassa, Jean-Baptiste Charbonnier, Jean-Claude Boulain, André Ménez, Roger Genet, and Muriel Gondry*

CEA/Saclay, Département d'Ingénierie et d'Etudes des Protéines, F-91191 Gif-sur-Yvette Cedex, France

Received July 22, 2005

Many studies that aim to characterize the proteome structurally or functionally require the production of pure protein in a high-throughput format. We have developed a fast and flexible integrated system for cloning, protein expression in *Escherichia coli*, solubility screening and purification that can be completely automated in a 96-well microplate format. We used recombination cloning in custom-designed vectors including (i) a (His)₆ tag-encoding sequence, (ii) a variable solubilizing partner gene, (iii) the DNA sequence corresponding to the TEV protease cleavage site, (iv) the gene (or DNA fragment) of interest, (v) a suppressible amber stop codon, and (vi) an S-tag peptide-encoding sequence. First, conditions of bacterial culture in microplates (250 μ L) were optimized to obtain expression and solubility patterns identical to those obtained in a 1-L flask (100-mL culture). Such conditions enabled the screening of various parameters in addition to the fusion partners (*E. coli* strains, temperature, inducer...). Second, expression of fusion proteins in amber suppressor strains allowed quantification of soluble and insoluble proteins by fluorescence through the detection of the S-tag. This technique is faster and more sensitive than other commonly used methods (dot blots, Western blots, SDS-PAGE). The presence of the amber suppressor tRNA was shown to affect neither the expression pattern nor the solubility of the target proteins. Third, production of the most interesting soluble fusion proteins, as detected by our screening method, could be performed in nonsuppressor strains. After cleavage with the TEV protease, the target proteins were obtained in a native form with a unique additional N-terminal glycine.

Keywords: protein expression • high-throughput screening • fluorescence • amber suppressor • S-tag • recombination cloning • fusion proteins

Introduction

The flood of sequence information coming from whole genome sequencing projects has ushered in the age of proteomics. Among the different areas of proteomics, structural proteomics commonly called structural genomics aims to characterize the repertoire of protein folds on a genome-wide scale.¹ One of the main bottlenecks to completion of this repertoire remains the ability to produce soluble and properly folded recombinant proteins or domains.²

Escherichia coli is the most extensively used organism for production of valuable gene products, as it generally provides sufficient protein material and offers many advantages: growth on inexpensive carbon sources, rapid biomass accumulation, well-characterized genetics, and an increasingly large number of mutant host strains,^{3,4} as well as easy labeling of proteins for NMR and crystallography. However, one disadvantage stems

from the frequent production of recombinant proteins as amorphous aggregates of misfolded molecules called inclusion bodies. To circumvent this problem, a universal approach does not exist but various and complementary strategies have been successfully proposed either to enhance *in vivo* solubility or to refold proteins trapped in inclusion bodies. Recently, some of these strategies were adapted or developed to be compatible with parallel processing of hundreds of proteins required for structural proteomics. In this way, screening of numerous buffers for *in vitro* refolding of proteins from inclusion bodies was adapted to a high-throughput format⁵ suitable for automation.⁶ An alternative route is offered by directed evolution in which protein diversity libraries are screened for improved solubility. This approach can reasonably be applied to structural proteomics thanks to new selection methods based on fusion solubility reporters⁷ like green fluorescent protein (GFP)^{8,9} or chloramphenicol acetyltransferase (CAT).¹⁰ However, although powerful, this approach remains time-consuming and has the potential drawback that soluble variants may be devoid of function.

The more conventional strategy of screening key parameters proved to be essential for enhancing *in vivo* solubility of wild-

* To whom correspondence should be addressed. Phone: + 33 169087647. Fax: + 33 169089071. E-mail: muriel.gondry@cea.fr.

[†] Both authors have equally contributed to this work.

[‡] Present address: University of Groningen, Center for Pharmacy, Department of Analytical Biochemistry, Antonius Deusinglaan 1, 9713AV Groningen, The Netherlands.

type proteins appears very attractive. Many parameters as fusion partners, *E. coli* strains, or induction temperatures, have been shown to affect protein solubility during the production phase.^{11,12} While optimization of these parameters is common for a given protein, it remains a major difficulty in high-throughput production of soluble proteins essentially due to the large number of parameters to be tested. To overcome the problem of numerous parameters, screening has to meet essential requirements as efficiency, rapidity, flexibility, and compatibility with the microplate format. Thus, each step in the screening process, i.e., subcloning, bacterial growth for expression and scoring of protein solubility, required adaptation or optimization. Prior to this work, although some steps had been separately miniaturized or optimized, no integrated screening approach had been reported that meets all the above criteria. These criteria were fulfilled for subcloning thanks to recent molecular methodologies, in particular ligation-independent cloning (LIC)¹³ and recombinational cloning.¹⁴ These methods have been successfully used in proteomics studies for subcloning a protein-encoding DNA into multiple expression vectors to produce this protein fused to various solubilizing partners.^{15,16} As regards the successive stages of protein expression including bacterial culture, cell lysis and separation of soluble proteins from insoluble ones, they have all been tentatively adapted to microscale processes using deep-well plates,^{17,18} but further investigations of culture volumes and cell lysis methods were still necessary. Finally, the main hurdle to overcome remained the high-throughput scoring of protein production and solubility. In previous studies, this step was performed by SDS-PAGE analyses¹⁶ or dot-blot procedures using hexahistidine-tagged proteins,^{19,15} which give only semi-quantitative results and break the use of easy handling liquid assay.

In the present paper, we describe a flexible and efficient pipeline for production of soluble proteins that is entirely carried out in standard microplates from subcloning to scoring of protein solubility. First, we used recombination cloning method in custom-designed vectors to produce target proteins fused to various hexahistidine-tagged solubilizing partners. Second, we optimized both bacterial growth and lysis in microtiter plates (250- μ L culture). Third, we screened expression conditions by fluorescence measurements through the complementation solubility tag-S assay.^{20,21} The S-tag-encoding sequence was set at the C-terminus of the target protein downstream of an amber suppressible codon: once the best solubility parameters are determined, native target proteins devoid of the solubility reporter S-tag can be produced without constructing new expression vectors simply by using a non-suppressor strain. We show that using the S-tag (i) does not modify the solubility profile, (ii) allows rapid determination of the best production conditions, and (iii) gives access to an estimation of the amount of native protein to be produced, making this method ideally adapted to high-throughput studies. Therefore, we have successfully developed a dual expression system suitable for structural proteomics as it can be used both for high-throughput screening of solubility parameters and production of native proteins.

Experimental Methods

Cloning. Target Selection. Target protein fragments correspond to domains predicted to fold independently and originating from three human proteins named in this study Hs1, Hs2, Hs3 (Hs used for *Homo sapiens*) and one yeast

protein named here Sc1 (Sc used for *Saccharomyces cerevisiae*) involved in the DNA damage response. The limits of these domains were identified from database searches for analogous sequences in different organisms (PSI-BLAST),²² by sequence comparison with analogous domains (MUSCLE),²³ and by secondary structure prediction (JPred).²⁴ To be selected, each protein fragment had to be analogous to sequences found in different primary structure contexts, and to exhibit at each extremity about 10 nonconserved hydrophilic residues containing no predicted secondary element. An additional human protein (named in this study Hs4) involved in DNA repair was targeted as a whole. This protein contains three identified structural domains linked by predicted unstructured regions.

This selection finally led to seven domains (named here D; the different domains of a protein are named Da, Db, ...) from proteins Hs1, Hs2, Hs3, and Sc1: respectively three domains from Hs1, Hs1Da (14.2 kDa), Hs1Db (13.6 kDa), Hs1Dc (19.3 kDa); two domains from Hs2, Hs2Da (16.1 kDa), Hs2Db (11.3 kDa); one domain from Hs3 and Sc1, Hs3Da (44.9 kDa), Sc1Da (27.8 kDa), and the full-length protein Hs4 (69 kDa).

Parallel Cloning of Target Sequences into Multiple Fusion Protein Expression Vectors. All expression vectors were constructed using the Gateway cloning technology (Invitrogen). First, the *attB*-flanked PCR products encoding the target proteins were amplified from corresponding DNA templates using two successive PCR reactions performed with the Expand High Fidelity PCR system (Roche) following the instructions of the manufacturer. The first PCR enabled the addition to the protein-encoding sequence of (i) the TEV cleavage site-encoding sequence to the 5' terminus, and (ii) the S-tag-encoding sequence to the 3' terminus (forward primer, 5'-GGCTTC-GAGAATCTTTATTTTCAGGGC(N)₂₄-3' and reverse primer 5'-TTAGCTGTCCATGTGTTGGCGTTCGAATTAGCAGCAGCGG-TTCTTTCTA(N)₂₄-3' where N is specific to the target sequence). The second PCR was performed using 10 ng of the first PCR product and primers containing the *attB* sequence (forward primer, 5'-GGGGGGGGGACAAGTTTGTACAAAAAAG-CAGGCTTCGAGAATCTTTATTTTCAGGGC-3' and reverse primer 5'-GGGGGGGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAC-TGTCCATGTGCTGGCGTTCGAA-3'). Second, the *attB*-flanked PCR products were recombined with the pDONR221 donor vector (Invitrogen) in BP clonase reaction to generate entry vectors. All entry vectors were sequenced between the two-recombination sites using ABI PRISM 310 Genetic Analyzer (Applied Biosystem). Entry vectors and destination vectors (detailed below) were used in LR clonase subcloning reaction to generate expression vectors. These recombination mixtures were used for transformation of DH5 α cells and positive clones were selected after analysis by colony PCR.

Design of Destination Vectors. Destination vector pDEST17 (Invitrogen) was used to express target proteins with an N-terminal (His)₆ tag. For purification facilities, the other destination vectors were modified in order to express a recombinant fusion protein with an N-terminal (His)₆ tag. The pDEST-His-Trx vector to express target proteins fused to tagged-thioredoxin was constructed from pDEST17 as follows: a *XhoI*-restriction site was inserted downstream of the (His)₆ tag-encoding sequence using the QuickChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The oligonucleotides used were 5'-CACCATCAC-CATCACCTCGAGTCAACAAGTTTGTAC-3' and the complementary primer. Thioredoxin gene was amplified from pBAD-DEST49 (Invitrogen) using the oligonucleotides 5'-CAT-

CACCTCGAGTCAGGATCTGATAAAATTATTCAT-3' and 5'-CTG-GTGGTCGACGGCCAGGTTAGCGTCGAGGAA-3'. After purification from agarose gel using the QIAquick Gel Extraction Kit (Qiagen), the PCR product was digested with *XhoI* and *Sall* and cloned into the *XhoI* site of pDEST-17-*XhoI*.

The pDEST15 (Invitrogen), pDEST-TH2, pDEST-TH3, and pDEST-TH5 vectors (kindly provided by T. Härd, KTH, Stockholm, Sweden) were modified as follows: a DNA fragment encoding the (His)₆ tag and obtained by hybridizing phosphorylated oligonucleotides 5'-TATGTCGTACTACCATCACCAT-CACCATCACCT-3' and 5'-TAAGGTGATGGTATGGTATGGT-TAGTACGACA-3' was cloned into the *NdeI* site of destination vectors. The resulting vectors were respectively named pDEST-His-GST, pDEST-His-NusA, pDEST-His-Gb1, pDEST-His-ZZ. All constructions were checked by DNA sequencing. Finally, the vector HispKM596 was obtained from C. Bignon (AFMB, Marseille, France) and enabled the expression of the protein of interest with an N-terminal (His)₆-tagged MBP fusion.

Cloning of Amber Suppressor tRNA Genes. Strains containing the amber suppressor tRNA genes, i.e., XA106 ($\Delta(gpt-lac)5$ *araFG gyrA metB argE_{am} rpoB thi-1 leuX151*) and DP50 *supF* (*fhuA53 dapD8 lacY1 glnV44* $\Delta(gal-uvrB)47$ *tyrT58 gyrA29 thyA57 hsdS3*) were obtained from Mary Berlyn (CGSC Stock Center).

The amber suppressor tRNA genes were first PCR amplified from *E. coli* bacterial strains XA106 for *leuX151* (*supP*) or DP50 *supF* for *tyrT58* (*supF*) and *glnV44* (*supE*) using Taq DNA polymerase (New England Biolabs). The oligonucleotides used were 5'-GTCGCATTGCATCTGATCGAATTCGGGGTATCGCCA-AGCGGTAAGG-3' and 5'-TTGAATCAAAATCTGGTCTGCAGTG-GCTGGGGTACGAGGATTCGAAC-3' for *glnV44*, 5'-GTCCGATAC-TACTGATGGAATTCGCCGAAGTGG-3' and 5'-CTTGAATATA-AATCTCGTCTGCAGTGGTGCCGAA-3' for *leuX151*, 5'-GTC-CGATTCGATCTGATGGAATTCGGTGGGGTCCCGAGCGGCCAA-3' and 5'-CTTGAATCAGAATCTCGTCTGCAGTGGTGGTGGG-GAAGGATTCGAAC-3' for *tyrT58*. They were designed using Oligo version 6 software (MedProbe, Oslo). The PCR conditions were 5 min at 95 °C, followed by 35 cycles of 30 s at 92 °C, 30 s at 50 °C and 30 s at 72 °C except for *leuX151* for which hybridization temperature was 30 °C. The PCR products were purified from agarose gel and then digested with *EcoRI* and *PstI* (in bold type in oligonucleotide sequences) and cloned between the *EcoRI* and *PstI* sites of pSUF putting the suppressor tRNA genes under the control of the constitutive *Plpp_{weak}* promoter.²⁵ Sequences of pSUF-*glnV44*, pSUF-*leuX151*, and pSUF-*tyrT58* (respectively named pSupE, pSupP, and pSupF) were checked using the ABIPrism 310 Genetic Analyzer.

Furthermore, the *Plpp_{weak}-glnV44* construct was amplified from pSupE using oligonucleotides 5'-TGATTCTATAAGTCTA-GATCGCCCCATCAAAAAATATTCTCAACATAAA-3' and 5'-AAGCAATCGAATCCAGCAGTCTAGAGCCAAGCTTAAAAAATC-CTTAGC-3' with the Expand High Fidelity PCR System (Roche) following the instructions of the manufacturer. After purification from agarose gel, the PCR product was digested with *XbaI* (bold type in nucleotides) and cloned into the *XbaI* site of pRARE, isolated from Rosetta strain (Novagen),²⁶ resulting in pRARE-SupE. Sequence was verified using the ABIPrism 310 Genetic Analyzer.

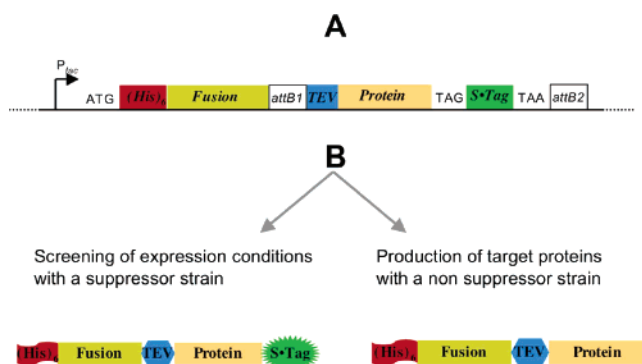
Protein Expression and Bacterial Cell Lysis. Different *E. coli* strains were used for expression: Rosetta (DE3) and Origami B (DE3) cells (Novagen), BL21 Star (DE3) and BL21-AI (Invitrogen), C41 (DE3), and C43 (DE3) (Avidis). Cultures were

induced with 1 mM IPTG except for cultures of BL21-AI cells which were induced with 0.2% L-arabinose.

Standard Large-Scale Process. The standard transformation was done as follows: chemically competent cells were transformed with 20 ng plasmid using standard heat-shock. After 1 h outgrowth without antibiotic, cells were plated on LB agar containing appropriate antibiotics and cultivated at 37 °C overnight. A starter culture was made by inoculating 3 mL LB containing the same antibiotics with about five fresh colonies and growing at 37 °C overnight. This starter culture was used to inoculate 100 mL of selective LB in a 1-L flask to OD₆₀₀ \approx 0.05. Two different expression protocols were performed: in the first one, expression culture was incubated at 37 °C to early-log phase (OD₆₀₀ \approx 0.5) and induced at 37 °C for 3 h; in the second one, induction was done at OD₆₀₀ \approx 1.2 and culture was continued at 20 °C for 16 h. Cells were harvested by centrifugation at 3000 \times g for 20 min, frozen at -80 °C, thawed and suspended in 5 mL lysis buffer L (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% glycerol, 10 μ M phosphoramidon, 1 mM PMSF), then lysed with an Eaton press (Rassant). The lysate was supplemented with 10 mM MgCl₂ and treated with 125 U benzonase (Merck) for 5 min at room temperature. The soluble protein fraction was separated from the insoluble fraction by centrifugation at 40 000 \times g for 30 min. The soluble proteins and the insoluble fraction (suspended in 5 mL 2% SDS) were further analyzed by SDS-PAGE (see below).

Microscale Process. Transformation in microplates (BD Falcon, 370- μ L useful volume) was done as above except that plating was replaced by addition of appropriate antibiotics directly in the 250 μ L transformation mixture. Microplates were sealed by a breathable sealing membrane (Nunc). Cells were grown at 37 °C overnight with 1000 rpm shaking (Thermomixer, Eppendorf). OD₆₀₀ were measured with a Microplate Reader (Bio-Tek Instruments) and cultures were diluted to a final OD₆₀₀ \approx 0.05 in 250 μ L of fresh LB containing the same antibiotics. Expression cultures were grown as previously described. Cells were harvested by centrifugation at 3000 \times g for 20 min (microplate carrier in Jouan 412), frozen at -80 °C, thawed and suspended in 50 μ L lysis buffer M (100 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X-100, pH 8.0) containing 0.2 mg/mL lysozyme (Sigma), 10 μ M phosphoramidon, and 1 mM PMSF. The 96-well microplates containing bacterial suspensions were kept at 4 °C with gentle shaking for 1 h. Each lysate was then treated with 1.25 U benzonase after addition of 10 mM MgCl₂, and microplates were incubated at room temperature for another 5 min. The soluble protein fractions were then separated from the insoluble fractions by centrifugation at 2830 \times g for 90 min. Soluble fractions were transferred to new 96-well microplates, and cell pellets were suspended in 50 μ L 2% SDS. Protein expression and solubility were further analyzed by SDS-PAGE and fluorescence assay (see below).

Analysis of Protein Expression and Solubility by SDS-PAGE. The samples for SDS-PAGE analysis were prepared in a microplate format by mixing 10 μ L of protein fraction (soluble or insoluble) and 10 μ L of an SDS-based denaturing sample buffer containing β -mercaptoethanol. Sealed microplates containing samples were incubated at 100 °C for 10 min and centrifuged. 8 μ L of samples and ladder (Precision plus Protein Standards from Biorad) were loaded into wells of 12% acrylamide gels. 12 SDS-PAGE gels were run simultaneously using the Dodecacell system of Biorad. Analysis of the expression and solubility patterns was performed after Coomassie blue stain-

Scheme 1. Dual Expression System

A, Design of expression vectors pEXP-X. The elements essential for translation are shown. All the vectors were obtained as follows: a 5'-(His)₆ tag-encoding sequence, a fusion partner gene, an *attB* recombination site, a TEV recognition site sequence, the target gene, a suppressible amber codon, an S-tag peptide-encoding sequence and finally a 3'-*attB* recombination site. **B**, Resulting fusion proteins. Screening of expression conditions was achieved with a suppressor strain: the fusion proteins were labeled with S-tag. Production of target proteins was performed with a nonsuppressor strain: the fusion proteins were devoid of S-tag.

ing. Relative intensities of the bands of interest were measured by densitometry (Bioprobe Systems).

Analysis of Protein Expression and Solubility by Fluorescence Assays. S-tag-labeled proteins produced in suppressor strains were quantified using the FRETworks S-tag Assay Kit (Novagen). Briefly, the soluble or insoluble protein fractions resulting from centrifugation after lysis were diluted 20- and 400-fold. 5 μ L of diluted samples were added to 195 μ L reaction mix in black nonbinding 96-well plates (Corning) and fluorescence was recorded for 10 min (492 nm excitation filter and 520 nm emission filter). Slopes from the linear part of the resulting fluorescence kinetics were compared to standard curves obtained for known concentrations of free S-tag (0 to 40 fmol) in order to quantify S-tag-labeled proteins in the bacterial soluble and insoluble fractions (in "equivalent S-tag"). Soluble protein yields (mg/L) were estimated from the equivalent-S-tag data assuming an average 30% rate of suppression.

Results

Design and Construction of Dual Expression Vectors. Our expression system, designed in the context of structural genomics, had to be suitable both for protein solubility high-throughput screening and efficient purification of native proteins. Thus, we constructed all expression vectors by including (i) a (His)₆ tag-encoding sequence followed by (ii) a variable fusion partner gene, (iii) a TEV recognition cleavage site-encoding sequence, (iv) the gene (or DNA fragment) to be expressed, (v) a suppressible amber stop codon (TAG), and finally (vi) an S-tag peptide-encoding sequence (Scheme 1A).

The amber stop codon constituted the key element of our dual expression system: the rationale was to use an amber codon suppressor tRNA gene to express a protein labeled with S-tag during the screening stage and to produce a protein devoid of S-tag in a related nonsuppressor strain during the production stage (Scheme 1B). S-tag, a 15-amino acid peptide, allowed detection of tagged proteins through its capacity to interact with S-protein and reconstitute RNase activity.²⁰ Moreover, monitoring enzymatic activity by a fluorescence assay was a rapid, very sensitive and high-throughput compat-

ible method to quantify recombinant proteins.²¹ The small size of S-tag offered a priori the further advantage of not modifying too much the solubility of its carrier protein.

We also fused various partners (i.e., GST, Trx, MBP, ZZ, Gb1, or NusA) to target proteins (or domains) chosen according to their capacity to enhance protein solubility *in vivo*.^{16,27,28,12} Moreover, we systematically introduced a (His)₆ tag at their N-terminus (Scheme 1B) which enables easy, efficient and homogeneous purification using immobilized-metal affinity chromatography, whatever the solubilizing partner. Finally, the TEV recognition site (Glu-Asn-Leu-Tyr-Phe-Gln*Gly)^{29,30} was introduced between the fusion partner and the target protein (Scheme 1B). The cleavage by the TEV protease occurs specifically and efficiently between the Gln and Gly residues to release the purified native target protein with a single additional N-terminal glycine.

Globally, our vectors were constructed with the technology used by Härd and co-workers.¹⁶ We conserved cloning by recombination using Gateway technology which proved to be rapid, efficient and directional and usable in different expression hosts. Seven different destination vectors were constructed *de novo* or modified from existing ones to produce the target proteins fused to different N-terminal (His)₆ tag-labeled fusion partners and the single (His)₆ tag as a reference. Two PCR amplification steps yielded each gene of interest flanked on its 5' terminus by an *attB* recombination site followed by the TEV recognition site-encoding DNA sequence and on its 3' terminus by the TAG codon, the S-tag-encoding DNA sequence and a second *attB* recombination site. These DNA sequences were then inserted by a first recombination reaction into the donor vector resulting in the entry vectors. The second recombination reaction was used to move the genes of interest from the entry vectors into our seven designed destination vectors, creating the desired expression vectors.

Optimization and Validation of the Microscale Procedure.

First of all, the standard large-scale process usually carried out in our laboratory to produce soluble protein fractions had to be transposed to a microplate-scale process, suitable for high-throughput screening of protein expression and solubility. To optimize our microscale procedure, we determined the conditions of microscale analysis that gave the same results than large-scale analysis. As a model, we used a protein commonly produced in our laboratory as a GST fusion protein using the pGEX plasmid (pGEX-41 from H. J. Worman, Columbia University, NY). In large-scale production, we obtained a good yield of partially soluble protein with BL21 Star (DE3) cultivated overnight at 20 °C. First, we confirmed that transformation by adding appropriate antibiotics directly to the transformation mix was identical to plating on selective agar medium (i.e., the transformation efficiencies were found similar). Second, cell growth for protein expression was tested with different culture volumes dispensed either in 96-well microtiter plates (0.25 mL per well) or 96-, 48-, and 24-deep well plates (1 or 2 mL per well), and incubated at various speed and rotation radii. For each induced culture, shaking conditions were identified that most closely allowed routine bacterial growth. We finally selected 96-well microtiter plates instead of common deep well plates for all steps of the microscale procedure: these plates contain smaller volumes (250 μ L) while maintaining sufficient protein material for downstream manipulations, and directly fit into the usual laboratory instrumentation (like microplate spectrophotometer). 250 μ L cultures in 96-well microtiter plates incubated with 1000 rpm shaking (Thermomixer, Eppendorf)

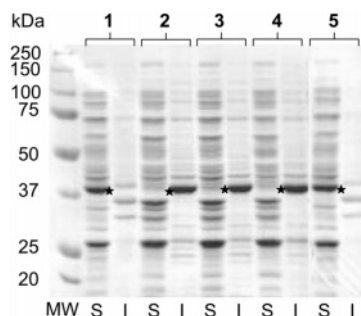


Figure 1. Determination of the best lysis conditions. SDS-PAGE analysis of soluble (S) and insoluble (I) protein fractions after bacterial lysis with an Eaton press (1), CellLytic B buffer from Sigma (2), B-Per buffer from Pierce (3), BugBuster HT buffer from Novagen (4), or our homemade buffer M (100 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X-100, pH 8) (5). Analysis was performed on a 12% acrylamide gel with Coomassie blue staining. The molecular weight ladder (MW, Precision Plus Protein Standards from Biorad) corresponded to the following weights (from top to bottom): 250 kDa, 150 kDa, 100 kDa, 75 kDa, 50 kDa, 37 kDa, 25 kDa, and 20 kDa. The band at around 26 kDa observed in all soluble fractions corresponded to the GST partner alone.

reached final OD₆₀₀ close to those obtained using the standard protocol (about 10–20% lower). Moreover, recombinant protein expression patterns were identical, as analyzed by SDS-PAGE (data not shown). Third, we determined optimal microplate lysis conditions. At the end of the induction cultures, cells were harvested by centrifugation with a microplate carrier. Cell pellets were then lysed using 4 different lysis buffers (*B-Per* (Pierce), *CellLytic B* (Sigma), *BugBuster HT* (Novagen), and a laboratory-made buffer M, detailed in Experimental Methods) to determine which one was giving the results closest to those of reference mechanical lysis in terms of soluble/insoluble protein profiles. The corresponding SDS-PAGE gel is shown in Figure 1. This gel showed that soluble and insoluble protein fractions were efficiently separated by centrifugation (3000 × g for 90 min). It also showed that the different lysis buffers did not give identical soluble/insoluble protein distributions. Only the buffer M gave results similar to those obtained with the reference (Figure 1, profiles 1 and 5, respectively). The other commercial reagents showed a tendency to favor aggregation of the protein as the expressed fusion was detected in the insoluble fraction (Figure 1, profiles 2, 3, and 4). The procedure finally optimized is detailed in the Experimental Methods and was used for all experiments described hereafter.

The microplate procedure optimized for one protein then required validation in real conditions of screening. This validation was carried out on a series of 8 proteins (or domains) expressed in fusion with the 7 selected partners (see above), in 3 strains (BL21 Star (DE3), Rosetta (DE3) and BL21-AI), and at 2 temperatures (37 °C and 20 °C). For 3 target proteins (Hs3Da, Hs4, and Sc1Da), screening was incremented with 3 additional strains (Origami B (DE3), C41 (DE3), and C43 (DE3)). Protein expression and solubility were analyzed by SDS-PAGE analysis. The microscale process we developed was particularly suitable for managing the 462 resulting expression conditions, as it required the use of only 5 dedicated microplates and could be performed within 24 h. On the other hand, despite the use of the Dodecacyll system (Biorad) which enabled us to run 12 gels at once, SDS-PAGE analysis was particularly tedious and time-consuming (924 samples loaded onto 66 gels), emphasizing

the need for alternative methods for high-throughput approaches. For each target protein, the best expression condition was chosen to perform a scale-up in the standard large-scale protocol. For 5 proteins (Hs1Da, Hs1Db, Hs2Da, Hs3Da, and Hs4) out of 8, the standard protocol gave expression and solubility profiles identical to the profiles obtained in microplates. For the 3 other ones, the results were even better in terms of expression and solubility (data not shown). The results obtained on scale-up confirmed the suitability of the microplate protocol for solubility screening.

Expression and Screening in Suppressor Strains. Using tRNA suppressor genes is a tricky matter because of the changes they may induce in cellular physiology. We were therefore aware of the necessity to choose a suppressor tRNA gene that did not substantially modify bacterial physiology and we checked the effects of expressing different tRNA genes on bacterial transformation, growth and protein expression profiles.

Many synthetic and natural amber suppressors have been reported to have different suppression efficiencies. Three of them characterized by different ranges of suppression efficiency were initially selected for our experiments: *glnV44* or *supE* (1–26% suppression efficiency, inserting a glutamine), *tyrT58* or *supF* (11–100% suppression efficiency, inserting a tyrosine) and *leuX151* or *supP* (30–100% suppression efficiency, inserting a leucine).³¹ Each plasmid carrying a suppressor gene was introduced into BL21-AI cells. We then compared transformation efficiency of the resulting suppressor strains with expression plasmids as well as their growth rates. Under the experimental conditions used, only BL21-AI cells bearing pSupE gave satisfactory transformation rates. Suppression by *supE* (*glnV44*) was hence used for all the following experiments.

We had previously selected six *E. coli* strains for our expression and solubility screening. The suppressor feature was introduced in the four most interesting ones: Star (DE3) and Rosetta (DE3) strains, which gave the best yields of protein solubility in our screening tests; BL21-AI strain, which possesses a tight regulated arabinose-inducible promoter and C41 (DE3) strain selected for its ability to express toxic recombinant proteins.³² Plasmid pSupE was introduced into Star (DE3), BL21-AI and C41 (DE3) strains yielding the corresponding suppressor strains Star*, AI* and C41*, respectively. This approach was not possible for the Rosetta (DE3) strain because of the presence of the pRARE plasmid possessing the same replicon (P15A) and antibiotic resistance than pSupE. A new plasmid (pRARE-SupE) expressing both rare tRNAs and *supE* was constructed. Introducing pRARE-SupE into Tuner(DE3) cells yielded a new strain equivalent to a Rosetta expressing the suppressor feature, Rosetta*.

The first step in validating our approach was to check the effects of *supE* tRNA gene on the different cell culture parameters in our microscale screening conditions. We repeated the whole screening procedure for 8 selected target proteins in the four suppressor strains (7 fusion partners, 2 temperatures). The efficiency of transformation in microplates was identical in both unmodified and suppressor strains except for plasmids expressing proteins as (His)₆, GST and Trx fusions in Star* (lower efficiency). The presence of the suppressor plasmid also had no influence on bacterial growth (similar generation time and same bacterial density at the end of the culture).

In a second step, we compared the expression and solubility profiles of the target proteins in the absence or presence of the tRNA suppressor gene. Cultures and cellular fractionation

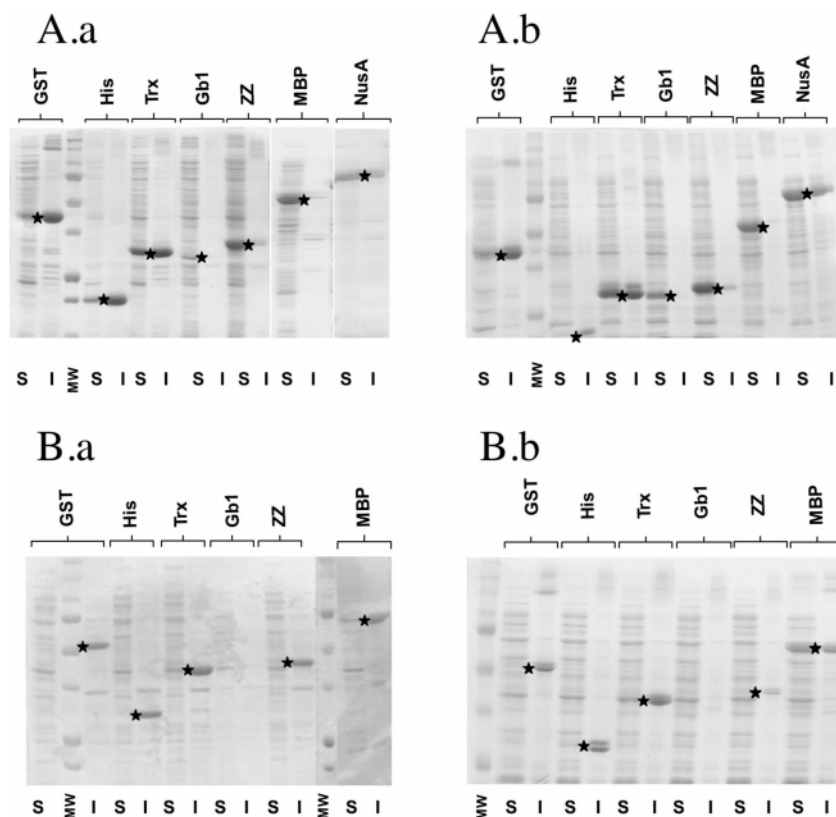


Figure 2. Comparison of expression and solubility patterns in nonsuppressor and suppressor *E. coli* strains. Soluble (S) and insoluble (I) protein fractions were analyzed by SDS-PAGE on 12% acrylamide gels and Coomassie blue staining. The proteins shown are Hs1Dc in Rosetta, at 20 °C (A) and Sc1Da in C41 at 37 °C (B), in the absence (a) and in the presence (b) of the suppressor plasmid. The molecular weight ladder corresponded to that described for Figure 1.

were performed according to our microscale procedure and samples were analyzed by SDS-PAGE. Figure 2 shows the results obtained for two proteins, Hs1Dc and Sc1Da, produced respectively in Rosetta cells grown at 20 °C and in C41 bacteria grown at 37 °C. Whatever the fusion partner, the target protein and the experimental conditions of culture used, the presence of the suppressor gene affected neither protein expression nor solubility (compare Figure 2a (*sup* −) and 2b (*sup* +)). Similar experiments were conducted for the other proteins and resulted in the same observations.

In a third step, it was essential to determine if protein suppression varied according to the fusion partner used. Large variations in suppression would bias the solubility quantification in the screening process. In most conditions, Coomassie blue staining of SDS-PAGE revealed two close bands of molecular weights expected for the proteins with and without the S-tag (as an example, see Figure 2A.b and 2B.b). The proportion of each protein reflected the relative efficiency of suppression. Each protein was quantified by densitometry (see below). Several conclusions could be drawn from this measurement. First, for a given protein, the suppression efficiency was identical whatever the fusion partner (and also the growth temperature). Although the suppression rate was not strictly identical in all the strains for one protein, the variation range was under 5% and could be considered as negligible. Second, the suppression rate could vary from one protein to another. For 6 proteins, the average suppression rate was around 30% (Hs1Da: $30 \pm 2\%$, Hs2Da: $29 \pm 2\%$, Hs2Db: $26 \pm 1\%$, Hs3Da: $26 \pm 3\%$, Hs4: $30 \pm 2\%$ and Sc1Da: $28 \pm 5\%$). For the last two, the rate dropped to around 15% (Hs1Db: $12 \pm 3\%$ and

Hs1Dc: $17 \pm 2\%$). This phenomenon is not surprising since suppression is known to be sequence-dependent. It did not introduce bias into the screening for a given protein but had to be kept in mind to interpret the quantification results obtained by fluorescence (as explained below, in the Discussion section).

Screening of Protein Expression and Solubility by Fluorescence Assays. Fluorescence assays were performed similarly on soluble (supernatant) and insoluble fractions (pellet solubilized in 2% SDS). The fractions were twice diluted to one-twentieth ($1/20$ and $1/400$) and $5 \mu\text{L}$ of these solutions were used to measure the amount of S-tag. The measurement was based on the complementation of the S-protein by the S-tag yielding fully active RNase. The re-formed RNase was then able to cleave a particular RNA substrate giving a fluorescent signal at 520 nm. The quantity of available S-tag was deduced from the comparison of the initial rate of RNA cleavage induced by our samples with a standard curve of cleavage rates by known concentrations of free S-tag (see Experimental Methods). The presence of the S-tag at the C-terminal part of the proteins enabled quantification of only the fully translated proteins. All the data were analyzed and processed using a short program created in Microsoft Excel. For each microplate well, the program indicated the initial rate of RNase reaction, the corresponding quantity of “equivalent S-tag”, the yield of soluble protein produced before and after cleavage by the TEV protease. To obtain the total yield of protein produced (with and without the S-tag), the program considered that the measurement obtained with the S-tag represented 30% of the whole protein produced (see previous paragraph).

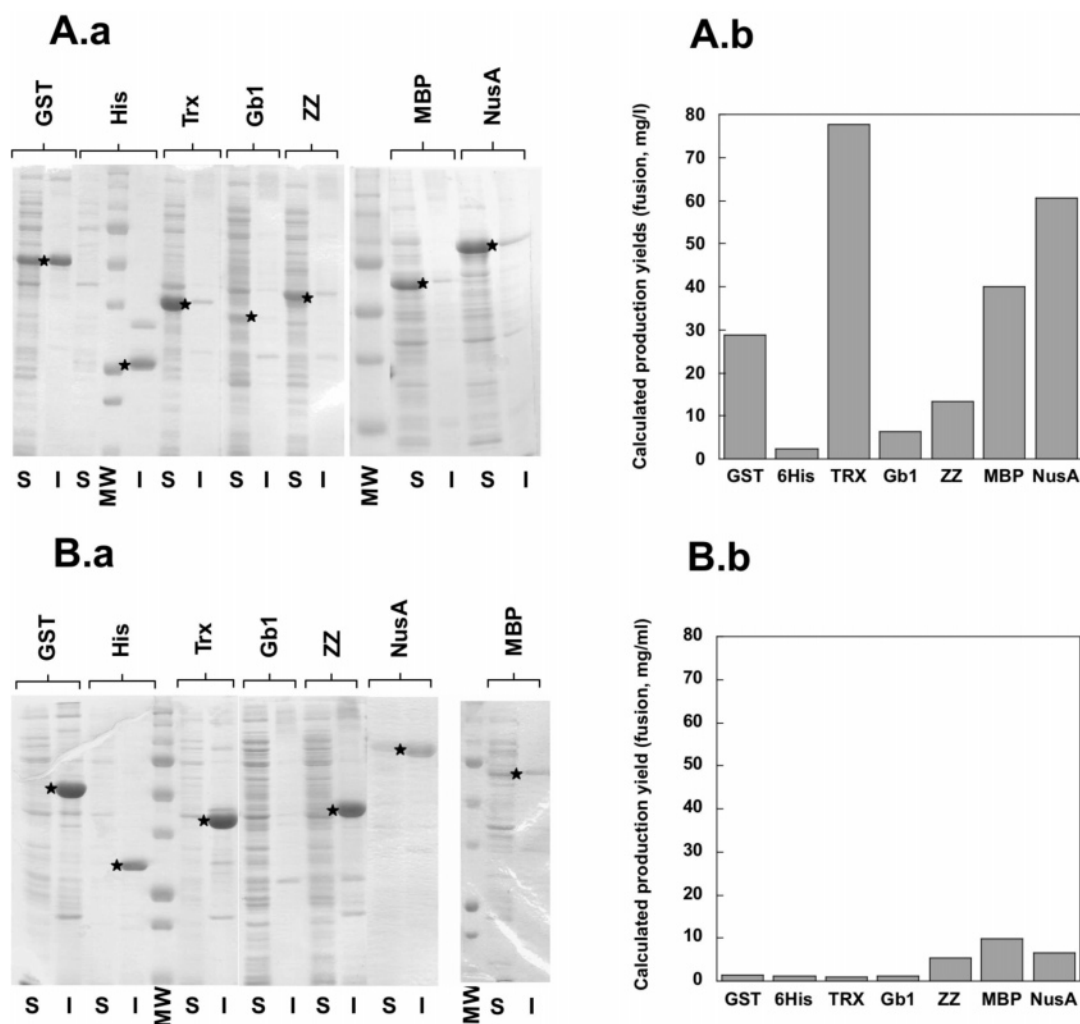


Figure 3. Comparison of the screening results obtained with soluble fractions by SDS-PAGE with Coomassie blue staining (a) and by S-tag detection (b). Production was performed in Rosetta* at 20 °C and soluble extracts were obtained as described in Materials and Methods. The proteins shown are Hs2Da (A) and Sc1Da (B). S-tag was quantified by fluorescence. Production yields (mg fusion protein/l) were calculated from S-tag amounts using an Excel macro and assuming an average 30% suppression rate. The molecular weights used for SDS-PAGE analysis are described for Figure 1.

The same culture samples were analyzed both by SDS-PAGE and fluorescence assays; thus, the results obtained from the two analysis methods could be strictly compared. It should be noted that the values obtained by fluorescence of insoluble fractions proved to be clearly underestimated (data not shown). The main problem to be solved to achieve relevant quantification of insoluble fractions was the solubilization of the pellets. 2% SDS was clearly insufficient to solubilize all aggregated proteins. No optimization of this step was performed in this study, as our goal was to enhance the *in vivo* solubility of the protein of interest. Hence, the following results essentially concern the soluble fractions.

Results obtained for all the proteins fused to different fusion partners and expressed in various strains at two distinct temperatures showed that the same solubility tendencies were obtained whatever the analysis method chosen. As an example, the Hs2Da domain observed by SDS-PAGE appeared more soluble (in descending order) when fused to Trx, NusA, MBP, and GST. The fluorescence data similarly indicated that Trx was the best fusion partner for production of soluble Hs2Da, followed by NusA, MBP, and GST (see Figure 3 A). Another example is given with the Sc1Da domain (Figure 3 B): only

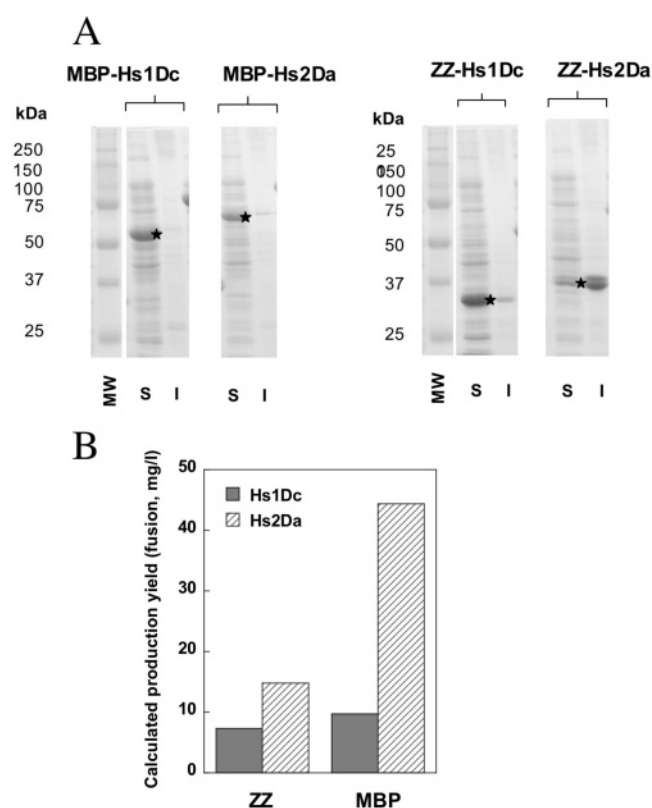
the fusion protein MBP–Sc1Da was easily detectable by gel analysis of soluble fractions and the best score of soluble fusion protein determined by fluorescence was indeed obtained with the MBP partner. This last example underlined the best sensitivity provided by fluorescence measurement compared to SDS-PAGE analysis; the soluble Sc1Da fusion proteins were not or hardly detectable on gels (the threshold of detection by Coomassie blue staining depends on the chemical composition of the target protein), but gave significant fluorescence signals (Figure 3B).

Another advantage afforded by S-tag measurement was the possibility of quantifying the protein produced (Table 1). When comparing these quantifications to the corresponding SDS-PAGE gels, the calculated yields appeared to be consistent with the intensity of the bands related to the proteins of interest for 6 proteins out of 8. For the 2 last proteins, Hs1Db and Hs1Dc, the calculated yields were clearly underestimated (see Figure 4 in which Hs2Da was an example of the results obtained with the 6 other proteins): similar intensities were obtained for both MBP–Hs1Dc and MBP–Hs2Da after Coomassie blue staining of SDS-PAGE gels (Figure 4A left), while the yields calculated from S-tag quantification differed by a factor 4

Table 1. Soluble Production Yields (mg/L of bacterial culture) Obtained by Fluorescence Assays for the 8 Target Proteins (or domains) Fused to Each of 7 Fusion Partners Cultivated at 20 °C in the Rosetta* Strain^a

	Hs1Da	Hs1Db	Hs1Dc	Hs2Da	Hs2Db	Sc1Da	Hs3Da	Hs4
(His) ₆ tag	41.36	3.36	5.49	2.63	3.43	1.26	2.90	12.25
	34.14	2.75	4.75	2.22	2.71	1.14	2.72	11.74
GST	105.26	3.63	3.16	32.07	5.23	1.70	5.53	12.39
	34.12	1.14	1.24	11.31	1.45	0.82	3.33	8.68
Trx	78.89	4.73	12.75	86.11	35.43	0.93	5.63	15.05
	38.23	6.04	7.15	44.51	15.23	0.61	4.21	12.35
Gb1	14.45	0.67	2.40	7.13	2.73	1.93	1.93	2.96
	8.31	0.38	1.55	4.32	1.41	0.87	1.56	2.57
ZZ	88.58	7.29	7.33	14.78	14.02	5.96	8.12	14.29
	37.10	2.97	3.63	6.66	5.13	3.50	5.65	11.12
MBP	91.87	6.56	9.76	44.39	24.98	10.89	18.75	16.42
	22.61	1.55	3.00	12.01	5.16	4.24	9.52	10.09
NusA	5.79	26.37	17.55	67.20	50.62	7.36	10.96	16.82
	1.14	5.02	4.38	14.64	8.29	2.39	4.78	9.17

^a The yields of fusion proteins (in bold) were extrapolated from S-tag amounts using an Excel macro and assuming an average 30% suppression rate. The expected yields of native proteins (in regular) were calculated considering a 100% cleavage efficiency.

**Figure 4.** Profiles obtained for proteins Hs1Dc and Hs2Da by SDS-PAGE after Coomassie blue staining (A) compared to estimation of the production yields from S-tag measurements (B). The gels correspond to the samples used for S-tag measurement and obtained in the presence of the suppressor plasmid (Rosetta*, 20 °C).

(Figure 4B). In the case of ZZ-fusions, the amount of soluble protein was clearly greater in the case of Hs1Dc compared to Hs2Da (Figure 4A right), whereas the yields extrapolated from S-tag measurement predicted twice as much soluble ZZ-Hs2Da (Figure 4B). Both Hs1Db and Hs1Dc have been found to have weaker suppression rates than the average value used in our calculations (15% instead of 30%). Nevertheless, the discrepancy between calculated yield and yields expected from SDS-PAGE gels remained significant even after correction for the suppression rates.

Finally, the possibility of calculating the expected quantity of protein that will be obtained after cleavage from its fusion partner also constituted an important piece of information. As an example, the results obtained for Hs1Da at 20 °C in Rosetta* indicated that GST was the best fusion partner (Table 1, extrapolated yields of fusion proteins in mg/L of culture are indicated in bold). In fact, after cleavage of the target protein from its fusion partner, the calculation gave very similar final yields of target protein for four different constructions (Table 1, figures in normal typeface). Hence, the choice of a fusion partner for expressing such a protein could be based on new additional criteria such as subsequent use of the protein or purification scheme.

Discussion

We have developed a fluorescence-based screening of soluble protein expression that enables testing of a large number of culture parameters and is suitable for high-throughput approaches as it can be completely automated. The whole approach consists of several steps, all performed in a 96-well microplate format, including the following: (i) PCR amplification and construction of expression vectors by recombination; (ii) transformation of suppressor strains of *E. coli* and culture in the chosen conditions; (iii) bacterial lysis and separation of the soluble protein fraction from the insoluble one; (iv) quantification of soluble (and possibly insoluble) proteins. Then, for each target protein, the expression vector leading to the best production of soluble protein can be directly introduced in an equivalent nonsuppressor *E. coli* strain giving a protein devoid of any tag after purification and cleavage by TEV protease (except an additional N-terminal glycine). Furthermore, the presence of a (His)₆ tag in the N-terminus of all fusion proteins allows an identical first purification whatever the fusion partner chosen, and so can potentially be automated.

The choice of a quantification method was crucial in our screening process. Direct detection methods that involve separating soluble proteins from insoluble ones and subsequent quantifying of these fractions by electrophoresis and/or immunological detection have been largely and commonly used.^{15–17,19} The major drawbacks of such methods are the difficulty of quantifying protein expression levels precisely as well as the fact that these methods are often time-consuming and labor-intensive. An attractive alternative for protein solubility screening consists of using fusion reporter methods. In

this way, green fluorescent protein (GFP) and chloramphenicol acetyltransferase (CAT) have been used to monitor protein expression and solubility.^{8,10} Although quite different in their principle of detection, these reporter proteins are large and can influence or disrupt the expression and solubility pattern of the proteins they are fused to. Thus, we focused on complementation solubility reporter assays that use small peptide tags fused to the target proteins: a small tag might be expected to have a limited effect on protein solubility. Two small reporters have been described recently: *lacZα* (β galactosidase α peptide), a 55-residue peptide (commercialized by Clontech under the name ProLable), complementing inactive *lacZΩ* (β galactosidase Ω peptide) to restore β galactosidase activity^{33,34} and S-tag, a 15-residue peptide complementing S-protein to reconstitute RNase S.^{20,35} Both of them enabled sensitive detection of labeled proteins and could be easily converted to a high-throughput format. We finally selected the S-tag peptide reporter because of its very small size. The 8 proteins tested in our series exhibited the same expression and solubility patterns in the presence and absence of S-tag. Nevertheless, the possibility that the presence of S-tag could have an influence must not be excluded, even if its small size and the nature of its amino acids made the probability very low. Indeed, it was recently reported that the presence of S-tag had a beneficial effect on the production yield of a sea anemone peptide, BgK.³⁶

Our dual expression system relies on the alternative use of related *E. coli* strains expressing or not an amber codon suppressor feature (during the screening and production steps respectively): expressing a tRNA suppressor gene results in the production of a recombinant protein with a C-terminal S-tag allowing quantitative detection in soluble extracts. Once more, the results obtained in the screening phase could be extended to a production stage only if the suppressor had no or limited influence on protein expression and solubility. In addition to tagging the protein of interest with the S-tag, the suppressor gene is also expected to affect any bacterial protein whose gene terminates with the suppressible stop codon. To reduce these effects, we chose to use an amber codon because it is the least common of the three stop codons, occurring at the end of 326 out of the 4,290 open reading frames in the *E. coli* genome.³⁷ Furthermore, amber suppression has been efficiently used for various applications such as phage display technology without showing significant modification of bacterial physiology. An amber codon was therefore introduced between our target genes and the S-tag-encoding sequence. Because the effects induced by expression of the suppressor were expected to be proportional to the efficiency of suppression, three different amber suppressor genes exhibiting different efficiencies were examined in our system. The use of the least efficient one, *supE* or *glnV44*, resulted in transformation efficiency and culture parameters similar to those obtained with the wild-type strain. Herring & Blattner recently examined global transcriptional effects of an amber suppressor and the few transcriptional effects observed appeared to be inconsequential for essential cell functions.³⁷ This is in favor of minimal modification of the protein synthesis machinery in amber suppressor strains compared to nonsuppressor ones. Our results reinforced this assumption. The analysis by SDS-PAGE and Coomassie blue staining of expression and solubility for 8 proteins showed the same patterns in the presence and absence of the suppressor. Whatever the strain considered, we observed that: first, the total amount of target protein (fused or not to S-tag) in suppressor strains is similar to the amount obtained in

equivalent nonsuppressor strains, and second, for a given protein, the amount of target protein fused to S-tag is proportional to the total amount whatever the N-terminal fusion partner. All these data demonstrated the suitability and significant reliability of our approach based on a screening performed in amber suppressor strains for identification of conditions for soluble expression of a target protein.

Although it does not directly interfere with the determination of the best expression conditions for a given protein, we observed a variation in suppression efficiency for some proteins. The suppression rate was around 30% for most of the proteins but two of them exhibited significantly lower rates. Several parameters are known to affect suppression efficiency. It is the result of the competition between elongation (read-through) mediated by the suppressor tRNA and termination mediated by release factors (RF1 in the case of amber codons). This efficiency is sensitive to both the 5' and 3' sequence context of stop codons, and *supE* tRNA is especially sensitive.^{31,38–41} The UAG decoding efficiencies of an amber suppressor tRNA on one hand and RF1 on the other hand are in particular dependent on the 3' flanking nucleotide: an A nucleotide causes a high read-through while RF1 interacts more readily with UAG if uracil is the 3' base.⁴¹ In our case, the 3' base is identical in all the constructions and was chosen to promote elongation (AAA codon). Concerning the influence of 5' sequence on suppression efficiency, read-through is not related to any properties of the mRNA sequence itself in 5'. Instead, it is the amino acids at the C-terminal end of the nascent peptide that influence termination. This is in agreement with our findings as we observed similar suppression rates for a given protein whatever the N-terminal fusion partner (same flanking sequence of stop codon in 3' and 5') and variable rates between target proteins (variable sequences in 5' of the TAG). Isaksson and co-workers⁴⁰ demonstrated that the 2 last amino acids in particular influenced termination efficiency. For the penultimate residue (−2 location), charge and hydrophobicity are important. Thus acidic residues such as aspartic and glutamic acids give inefficient termination (high read-through) and basic residues such as arginine and lysine promote efficient termination.⁴² For the last amino acid (−1 location), secondary structure propensities are determining factors and favor termination.⁴³ Among all our proteins tested, only 2, Hs1Db and Hs1Dc, exhibited significantly lower suppression rates than the others (15% versus 30%). Both Hs1Db (100 residues) and Hs1Dc (160 residues) are domains derived from the same protein, Hs1Db being the C-terminal part of Hs1Dc. The C-terminus peptide is thus identical, which is coherent with similar suppression efficiency. The final C-terminal dipeptide comprises 2 lysine residues, which should effectively favor termination against suppression according to the results obtained by Isaksson and co-workers.⁴² Although this effect has no influence on the determination of the best expression conditions for the proteins concerned, it entails an underestimation of protein yields by quantification with S-tag. Our calculation program considers as routine an average value of 30% for the suppression rate. This inconvenience can easily be overcome in cases as ours where protein domains are concerned. The nature of the 2 last residues must be taken into account in the delimitation of the domain and, if necessary, even modified to meet the criteria for suppression efficiency. In the case of full proteins, the estimation of production yields must be considered in the light of the last dipeptide sequence and its potential influence on suppression rate.

Quantification of protein yields by measurement of S-tag amounts proved to be both very sensitive and compatible with an automated large-scale screening. The predicted yields proved to be a good estimation of the yields obtained in scale-up for most of the proteins of our series. As discussed above, the variation in suppression rates explained some of the discrepancy between predicted and obtained yields. Nevertheless, in the case of Hs1Db and Hs1Dc, even after correction for the suppression rate, the predicted yields were still significantly lower than what could be expected from SDS-PAGE analysis, and then the yields actually obtained after scale-up. The second factor that probably also plays an important part in the underestimation of the yields is the accessibility of the S-tag peptide. Fluorescence measurement relies on the complementation of S-protein by S-tag to reconstitute a functional RNase. S-tag, therefore, has to be sufficiently accessible and flexible to adopt its native helical conformation. These features are difficult to predict and probably depend on the protein carrier. In our constructions, we introduced no linker between the protein of interest and the S-tag except the amino acid encoded by the amber codon. This may allow insufficient flexibility in some cases. An additional denaturation step before S-tag quantification could make the tag more accessible and lead to more accurate quantification. However, such an additional step was found to be time-consuming for little improvement in the whole process.

In conclusion, we have implemented a dual protein expression system that enables both fluorescence-based protein expression and solubility screening in high-throughput format and easy, potentially automatable production on the scale required by the experimenter. This process has since proven robust and flexible in the screening in our lab of production conditions for more than 50 proteins or domains. Screening could be adjusted to a large set of parameters ranging from particular *E. coli* strains, fusion partners to physical parameters of culture. In some cases, new parameters that prove beneficial for soluble protein expression, such as coexpression of chaperones or physiological partners, and medium additives,⁴⁴ could successfully be introduced in our screening without any modification of the overall process.

Acknowledgment. We thank M. Moreau for technical assistance during her professional training, S. Caputo, G. Charier, and A. Coirier (CEA, Saclay, France) for the large-scale production of native target proteins. We also thank Dr C. Bignon (AFMB, Marseille, France) for contributing the fusion vector HispKM596 and the group of Prof T. Härd (Royal Institute of Technology, Stockholm, Sweden) for contributing the vectors pDEST-TH2, pDEST-TH3, and pDEST-TH5. We also thank M. Berlyn and staff for sending the *E. coli* strains XA106 and DP50 *supF*. We are indebted to W. McClain for the gift of plasmid pSUF. Finally, we thank J. Angulo, S. Boiteux, J. P. Radicella (CEA, Fontenay-aux-Roses, France), D. Jullien (The Wellcome Trust Centre for Cell Biology, Edinburgh) and J.-P. de Villartay (INSERM, Hôpital Necker, France) for providing the DNA templates used in this study.

References

- Christendat, D.; Yee, A.; Dharamsi, A.; Kluger, Y.; Savchenko, A.; Cort, J. R.; Booth, V.; Mackereth, C. D.; Saridakis, V.; Ekiel, I.; Kozlov, G.; Maxwell, K. L.; Wu, N.; McIntosh, L. P.; Gehring, K.; Kennedy, M. A.; Davidson, A. R.; Pai, E. F.; Gerstein, M.; Edwards, A. M.; Arrowsmith, C. H., Structural proteomics of an archaeon. *Nat. Struct. Biol.* **2000**, *7* (10), 903–909.
- Edwards, A. M.; Arrowsmith, C. H.; Christendat, D.; Dharamsi, A.; Friesen, J. D.; Greenblatt, J. F.; Vedadi, M., Protein production: feeding the crystallographers and NMR spectroscopists. *Nat. Struct. Biol.* **2000**, *7* Suppl, 970–972.
- Baneyx, F., Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **1999**, *10* (5), 411–421.
- Baneyx, F.; Mujacic, M., Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **2004**, *22* (11), 1399–1408.
- Tresaugues, L.; Collinet, B.; Minard, P.; Henckes, G.; Aufrere, R.; Blondeau, K.; Liger, D.; Zhou, C. Z.; Janin, J.; Van Tilbeurgh, H.; Quevillon-Cheruel, S., Refolding strategies from inclusion bodies in a structural genomics project. *J. Struct. Funct. Genomics* **2004**, *5* (3), 195–204.
- Vincentelli, R.; Canaan, S.; Campanacci, V.; Valencia, C.; Maurin, D.; Frassinetti, F.; Scappucini-Calvo, L.; Bourne, Y.; Cambillau, C.; Bignon, C., High-throughput automated refolding screening of inclusion bodies. *Protein Sci.* **2004**, *13* (10), 2782–2792.
- Waldo, G. S., Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **2003**, *7* (1), 33–38.
- Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C., Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **1999**, *17* (7), 691–695.
- Pedelacq, J. D.; Piltch, E.; Liong, E. C.; Berendzen, J.; Kim, C. Y.; Rho, B. S.; Park, M. S.; Terwilliger, T. C.; Waldo, G. S., Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* **2002**, *20* (9), 927–932.
- Maxwell, K. L.; Mittermaier, A. K.; Forman-Kay, J. D.; Davidson, A. R., A simple in vivo assay for increased protein solubility. *Protein Sci.* **1999**, *8* (9), 1908–1911.
- Makrides, S. C., Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **1996**, *60* (3), 512–538.
- Sorensen, H. P.; Mortensen, K. K., Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **2005**, *4* (1), 1.
- Aslanidis, C.; de Jong, P. J., Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **1990**, *18* (20), 6069–6074.
- Walhout, A. J.; Temple, G. F.; Brasch, M. A.; Hartley, J. L.; Lorson, M. A.; van den Heuvel, S.; Vidal, M., GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **2000**, *328*, 575–592.
- Doyle, S. A.; Murphy, M. B.; Massi, J. M.; Richardson, P. M., High-throughput proteomics: a flexible and efficient pipeline for protein production. *J. Proteome Res.* **2002**, *1* (6), 531–536.
- Hammarström, M.; Hellgren, N.; van Den Berg, S.; Berglund, H.; Härd, T., Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **2002**, *11* (2), 313–321.
- Knaust, R. K.; Nordlund, P., Screening for soluble expression of recombinant proteins in a 96-well format. *Anal. Biochem.* **2001**, *297* (1), 79–85.
- Shih, Y. P.; Kung, W. M.; Chen, J. C.; Yeh, C. H.; Wang, A. H.; Wang, T. F., High-throughput screening of soluble recombinant proteins. *Protein Sci.* **2002**, *11* (7), 1714–1719.
- Vincentelli, R.; Bignon, C.; Gruez, A.; Canaan, S.; Sulzenbacher, G.; Tegoni, M.; Campanacci, V.; Cambillau, C., Medium-scale structural genomics: strategies for protein expression and crystallization. *Acc. Chem. Res.* **2003**, *36* (3), 165–172.
- Raines, R. T.; McCormick, M.; Van Oosbree, T. R.; Mierendorf, R. C., The S. Tag fusion system for protein purification. *Methods Enzymol.* **2000**, *326*, 362–376.
- Grabski, A.; Drott, D.; Mehler, M., Automated solubility screening of recombinant proteins in a 96-well format. *inNovations* **2003**, *16*, 11–13.
- Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402.
- Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792–1797.
- Cuff, J. A.; Clamp, M. E.; Siddiqui, A. S.; Finlay, M.; Barton, G. J., JPred: a consensus secondary structure prediction server. *Bioinformatics* **1998**, *14* (10), 892–893.
- Gabriel, K.; McClain, W. H., A set of plasmids constitutively producing different RNA levels in *Escherichia coli*. *J. Mol. Biol.* **1999**, *290* (2), 385–389.
- Novy, R.; Drott, D.; Yaeger, K.; Mierendorf, R., Overcoming the codon bias of *E. coli* for enhanced protein expression. *inNovations* **2001**, (12), 1–3.

- (27) Braun, P.; Hu, Y.; Shen, B.; Halleck, A.; Koundinya, M.; Harlow, E.; LaBaer, J., Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99 (5), 2654–2659.
- (28) Kapust, R. B.; Waugh, D. S., *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **1999**, 8 (8), 1668–1674.
- (29) Carrington, J. C.; Dougherty, W. G., A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85 (10), 3391–3395.
- (30) Kapust, R. B.; Routzahn, K. M.; Waugh, D. S., Processive degradation of nascent polypeptides, triggered by tandem AGA codons, limits the accumulation of recombinant tobacco etch virus protease in *Escherichia coli* BL21(DE3). *Protein Expr. Purif.* **2002**, 24 (1), 61–70.
- (31) Kleina, L. G.; Masson, J. M.; Normanly, J.; Abelson, J.; Miller, J. H., Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* **1990**, 213 (4), 705–717.
- (32) Miroux, B.; Walker, J. E., Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **1996**, 260 (3), 289–298.
- (33) Wigley, W. C.; Stidham, R. D.; Smith, N. M.; Hunt, J. F.; Thomas, P. J., Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat. Biotechnol.* **2001**, 19 (2), 131–136.
- (34) Eglen, R. M.; Singh, R., Beta galactosidase enzyme fragment complementation as a novel technology for high throughput screening. *Comb. Chem. High Throughput Screen* **2003**, 6 (4), 381–387.
- (35) Kelemen, B. R.; Klink, T. A.; Behlke, M. A.; Eubanks, S. R.; Leland, P. A.; Raines, R. T., Hypersensitive substrate for ribonucleases. *Nucleic Acids Res.* **1999**, 27 (18), 3696–3701.
- (36) Braud, S.; Belin, P.; Dassa, J.; Pardo, L.; Mourier, G.; Caruana, A.; Priest, B. T.; Dulski, P.; Garcia, M. L.; Menez, A.; Boulain, J. C.; Gasparini, S., BgK, a disulfide-containing sea anemone toxin blocking K⁺ channels, can be produced in *Escherichia coli* cytoplasm as a functional tagged protein. *Protein Expr. Purif.* **2004**, 38 (1), 69–78.
- (37) Herring, C. D.; Blattner, F. R., Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frmR. *J. Bacteriol.* **2004**, 186 (20), 6714–6720.
- (38) Bossi, L.; Ruth, J. R., The influence of codon context on genetic code translation. *Nature* **1980**, 286 (5769), 123–127.
- (39) Bossi, L., Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J. Mol. Biol.* **1983**, 162A (1), 73–87.
- (40) Mottagui-Tabar, S.; Isaksson, L. A., Only the last amino acids in the nascent peptide influence translation termination in *Escherichia coli* genes. *FEBS Lett.* **1997**, 414 (1), 165–170.
- (41) Pedersen, W. T.; Curran, J. F., Effects of the nucleotide 3' to an amber codon on ribosomal selection rates of suppressor tRNA and release factor-1. *J. Mol. Biol.* **1991**, 219 (2), 231–241.
- (42) Mottagui-Tabar, S.; Bjornsson, A.; Isaksson, L. A., The second to last amino acid in the nascent peptide as a codon context determinant. *Embo. J.* **1994**, 13 (1), 249–257.
- (43) Bjornsson, A.; Mottagui-Tabar, S.; Isaksson, L. A., Structure of the C-terminal end of the nascent peptide influences translation termination. *Embo. J.* **1996**, 15 (7), 1696–1704.
- (44) Schaffner, J.; Winter, J.; Rudolph, R.; Schwarz, E., Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins. *Appl Environ Microbiol.* **2001**, 67 (9), 3994–4000.

PR050230I