Medium-Scale Structural Genomics: Strategies for Protein Expression and Crystallization

RENAUD VINCENTELLI, CHRISTOPHE BIGNON,* ARNAUD GRUEZ, STEPHANE CANAAN, GERLIND SULZENBACHER, MARIELLA TEGONI, VALÉRIE CAMPANACCI, AND CHRISTIAN CAMBILLAU*

AFMB, UMR 6098, CNRS & Universités Aix-Marseille I & II, 31 Chemin J. Aiguier, 13402 Marseille Cedex 20, France

Received May 20, 2002

ABSTRACT

While high-throughput methods of protein production and crystallization are beginning to be well documented, owing to the output of large structural genomics programs, medium-throughput methods at the laboratory scale lag behind. In this paper, we report a possible way for an academic laboratory to adapt high-throughput to medium-throughput methods, on the basis of the first results of two projects aimed at solving the 3D structures of *Escherichia coli* and *Mycobacterium tuberculosis* (Tb) proteins of unknown function. We have developed sequential and iterative procedures as well as new technical processes for these programs. Our results clearly demonstrate the value of this medium-throughput approach. For instance, in the first 14 months of the *E. coli* program, 69 out of 108 target genes led to soluble proteins, 36 were brought to crystallization, and 28 yielded crystals; among the latter, 13 led to usable data sets and 9 to structures. These results, still incomplete, might help in planning future directions of expression and crystallization of proteins applied to medium-throughput structural genomics programs.

Introduction

Schematically, post-genomics comprises transcriptome, proteomics, and structural genomics (SG), which are the natural offspring of genomics. At the end of the 20th century, whole genome sequencing projects reached maturity, in terms of both results and technology. Today, sequencing of small bacterial genomes takes only a few days, and hence genomic information is continuously growing to feed these three new areas of research. According to the above definition, one would expect SG programs to solve the 3D structures of the whole proteome encoded by a given genome.[1−3] Unfortunately, even when a genome can be sequenced, only part of it can be exogenously expressed as soluble proteins. Furthermore, only a fraction of the soluble proteins produce crystals, at least at the present state of technology. Consequently, *actual* SG projects are in fact targeted projects. This led us to the conclusion that SG is not restricted to large consortia or companies but could be managed by academic laboratories as well. Nevertheless, the main difference between a large ($>n \times 1000$ genes) and a medium ($n \times 100$ genes) SG project remains a quantitative difference. Perhaps surprisingly, this has dramatic consequences on how the project is carried out. In a large project, most of the investment is made in the cloning and crystalization steps.[4] This is done at the expense of the intermediate stages, which generally involve a single set of expression conditions and a single protein purification step.[5] As shown in the present paper, the exact opposite situation specifies a medium-scale project. In particular, a target is not rejected if it does not respond to the first set of expression conditions, but rather it is tested under different experimental conditions.

We are currently involved in four medium-size SG projects, all of them related to human health and involving targets as dissimilar as proteins of unknown function from *Escherichia coli* (ASG) and *Mycobacterium tuberculosis* (Tb), mammalian membrane proteins, and enzymes of viral replication (see http://afmb.cnrs-mrs.fr/stgen/). Owing to this large diversity, these proteins demand widely divergent expression and crystallization conditions. This forced us to devise a flowchart made of several sets of conditions of increasing complexity (which we call "screening rounds") to span as many of these demands as possible. For instance, the screening data of ASG and Tb projects reported herein reveal that these two projects used different combinations of screening rounds, thereby suggesting that a given combination could be project-specific.

Most SG projects address medical issues in fields such as oncology, neurological, and infectious diseases that monogenic approaches failed to solve because the underlying molecular mecanism implicated more than one gene. In this respect, ASG and Tb projects are particularly representative of the stakes involved in SG. Both are collaborative projects. The first one, with a private company searching for new therapeutic targets, contains 108 *E. coli* open-reading frames (ORFs). The second one, with the Pasteur Institute, concerns 182 *M. tuberculosis* ORFs.

* Corresponding authors. Tel.: +33-491-16-45-01. Fax: +33-491-16-45-36. E-mail: cambillau@afmb.cnrs-mrs.fr, bignon@afmb.cnrs-mrs.fr.

Renaud Vincentelli is a CNRS engineer, dedicated to the development of new methods in protein expression.

Christophe Bignon is a CNRS engineer in charge of the robotized platform for protein expression.

Arnaud Gruez is a postdoctoral researcher in charge of the crystallization (nanodrops) and crystallography for structural genomics projects.

Stephane Canaan is a postdoctoral researcher in charge of the *Mycobacterium tuberculosis* structural genomics project.

Gerlind Sulzenbacher is a postdoctoral researcher in charge of the crystallization and crystallography for structural genomics projects.

Mariella Tegoni is a senior scientist who has developed the activity tests for redox proteins within the *E. coli* structural genomics project.

Valérie Campanacci is a postdoctoral researcher in charge of the *E. coli* structural genomics project.

Christian Cambillau is head of the AFMB laboratory at CNRS and is in charge of the structural genomics programs.

10.1021/ar010130s CCC: $25.00
Published on Web 01/04/2003

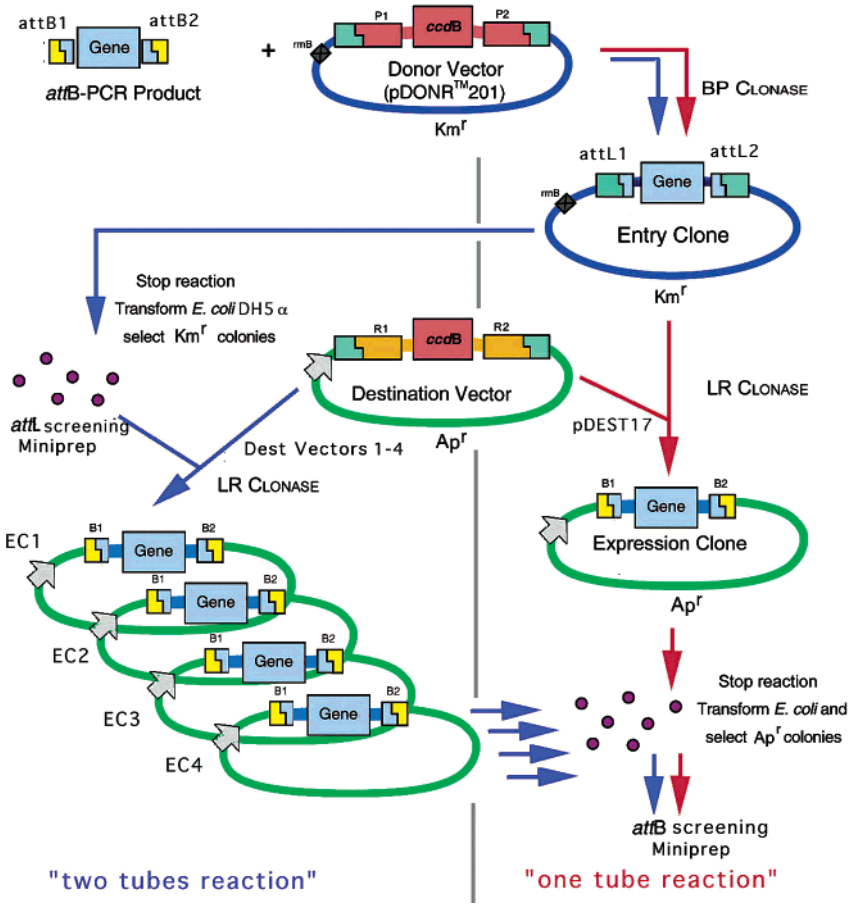

FIGURE 1. Recombination cloning: the Gateway technology. The ORF of interest was PCR amplified using either the whole *E. coli* genome or a cosmid bearing a fragment of the Tb genome as a template, and primers containing at their 5′ end the 5′ (attB1) and 3′ (attB2) recombination sites, respectively. The PCR product was then subcloned into a shuttle vector (pDONR201) by incubating for at least 1 h at 25 °C in the presence of the BP clonase enzyme. At this point, two pathways could be followed, depending on the number of expression plasmids to be used. *The one-tube reaction* (one expression vector, rounds 1 and 2): In this case, the ORF was directly transferred from the shuttle vector to the pDEST17 destination vector by incubating the BP reaction mixture for 1 h at 25 °C with LR clonase. DH5α cells were transformed with the whole mixture, and recombinant clones were selected by plating on ampicillin plates. Although the subcloning efficiency was close to 100%, for safety reasons we screen by PCR 2 colonies using destination vector-specific primers (attB1 and 2). *The two-tube reaction* (more than one expression vector, round 3): DH5α cells were transformed with the BP reaction mixture, and recombinant clones (so-called "entry clones") were selected by plating on kanamycin plates. Two colonies were screened using entry clone-specific primers (attL1 and -2). The intermediate construct was purified by miniprep from one positive clone. The ORF was then individually transferred from the entry clone to as many destination vectors as required using the LR reaction (see above). In ASG and Tb programs, four expression vectors have been used (labeled in the figure EC1−EC4). The efficiency of the system greatly relied on the fact that each recombination reaction substituted a lethal gene cassette with the ORF. Therefore, *E. coli* cells uptaking unrecombined vector would die, hence the very low background.

The first project started in June 2001 and the second in January 2003. The first ASG structures and Tb crystals are now available (January 2002).

Gene Cloning and Protein Expression

Processing several hundreds of genes simultaneously required a high-efficiency cloning strategy. The Gateway[6] technology (Invitrogen) was nearly 100% efficient. In addition, there was no restriction site analysis nor digestion of the target gene or of the vector prior to cloning, which hence saved time. A detailed description of this technology can be found in Figure 1 and at http://www.invitrogen.com.

For the reasons mentioned in the Introduction, we also devised an iterative strategy made of four rounds of screening of increasing complexity (Figure 2). In the first round, cloning and expression were performed with a basic set of conditions. The underlying reason was that it has been reported that up to 20% of the targets of a given SG project could produce soluble proteins and good crystals under very basic experimental conditions (the so-called "low-hanging fruits").[4] At the end of the first round, the remaining ORFs entered the second round, and so on. Today, ORFs refractory to all four rounds are abandoned (Figure 2). To assess the efficiency of the first round, 20 randomly chosen ASG and Tb targets were tested in a first trial. As expected, a significant fraction of the *E. coli* proteins expressed were soluble. By contrast, all Tb proteins yielded insoluble proteins. It was therefore decided that the Tb project would start directly at screening round 3. For that reason, *E. coli* targets were essentially

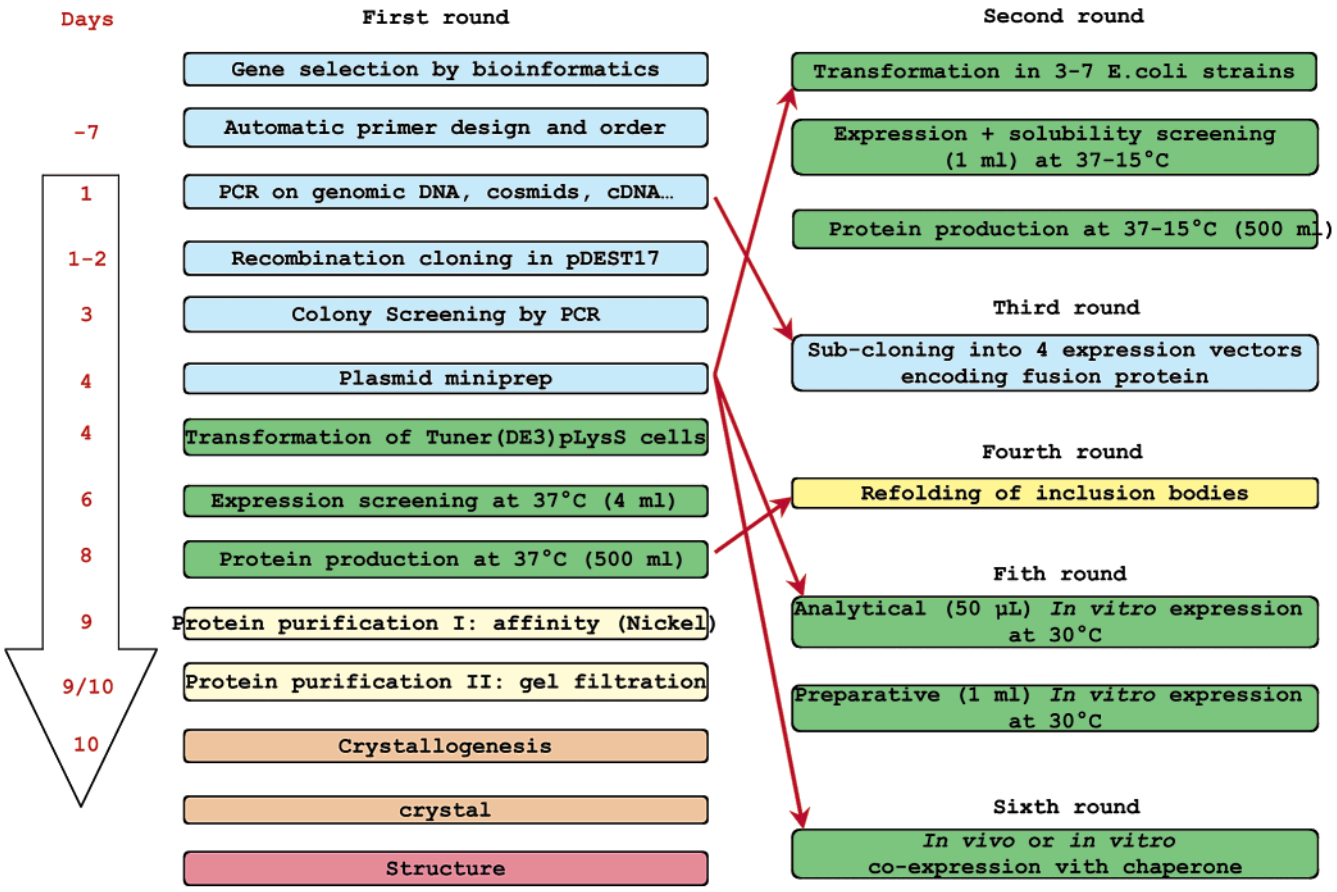

FIGURE 2. Evolution of the work scheme. The 108 target genes were processed in a simple first screening round. The targets that led to negative results (i.e., no expression or insoluble protein) underwent a second screening round, and so on. To date, rounds 1−4 but not 5 and 6 have been effectively tried in SG programs.

processed through rounds 1 and 2 and Tb targets through rounds 3 and 4.

First Round. We have defined a flowchart (Figure 2, first round) allowing two technical assistants using standard laboratory equipment to process 96 targets from the initial PCR to the crystal screening in about 10 days. In rounds 1 and 2, pDEST17 was used. This expression vector encoded an His6 tag linked to the N-terminus of the target protein by means of 12 residues encoded by the 5′ recombination site. The resulting fusion protein therefore bore 21 non-native amino acids at its N-terminus. We decided not to remove this tail in the first round to assess its possible deleterious effect on crystallization. Since a single expression vector was used, the cloning followed the "one-tube reaction" protocol (Figure 1, red arrows).

Protein expression was analyzed on 4-mL cultures, and, if positive, purification was directly performed on a larger volume. Tests were first performed in the BL21(DE3) strain, in which expression proved constitutive. We therefore switched to tuner(DE3)pLysS strain, in which expression is more tightly controlled.

In three weeks, most of the 108 targets of the *E. coli* project were cloned, and some were sent to crystallization trials. At the end of round 1 (3 months), nine proteins (8%) could be directly submitted to crystallogenesis trials, thereby confirming our survey of the literature in this respect. Six proteins gave crystalline material and two diffraction data sets (Figure 3, first round). Incidentally, this round provided evidence that crystallization was possible despite the extra N-terminal tail.

Second Round. Interestingly, BL21(DE3) and tuner(DE3)pLysS diverged not only in expression control but also in protein yield and solubility. This prompted us to extend the number of strains to test from 2 to 7. We also tried three incubation temperatures (37, 20, and 15 °C), as decreasing temperature is known to improve protein solubility. These conditions defined the second screening round. The seven *E. coli* strains were BL21(DE3), BL21(DE3)pLysS, Tuner(DE3)pLysS, OrigamiB(DE3)pLysS, Rosetta(DE3)pLysS, C41(DE3)[7] and C43(DE3).[7] The first five were obtained from Novagen, and the last two were from Avidis SA. They had different phenotypes but were used here empirically.

Combining temperatures and strains meant that each ORF had to be tested under $(7 \times 3) = 21$ expression conditions with $(2 \times 21) = 42$ gel loadings for analysis (to assess the soluble/insoluble protein ratio), resulting in 2646 gel loadings for the 63 remaining genes. We therefore decided to validate the second round with only 11 genes (462 experimental points) instead of 63 and to substitute sodium docecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) with a quantitative dot-blot procedure (Figure 4; manuscript in preparation).

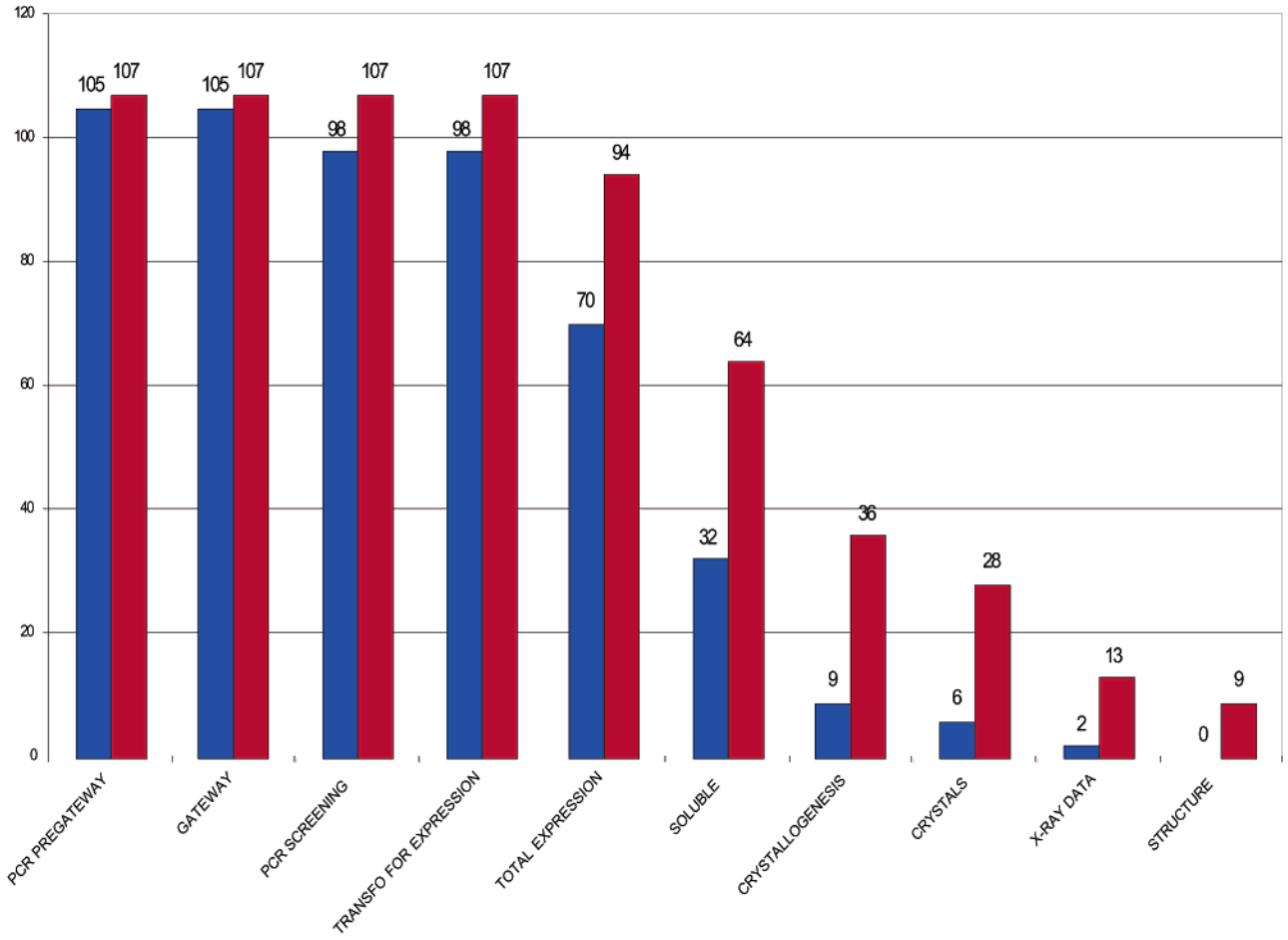

FIGURE 3. Step-by-step summary of the results of the first (blue) and second (red) rounds of screening. The different steps are depicted along the horizontal axis, in a chronological order from left to right. The absolute number of successful targets for a given step is indicated along the vertical axis. The exact number is written on top of each bar.

Five soluble proteins were obtained out of 11 genes. Cell cultures performed at 20 °C did not give better results than those performed at either 37 or 15 °C, and so this condition was skipped. Finally, three strains (BL21(DE3)-pLys, C41 and Origami(DE3)pLysS) turned out to be enough to assay for solubility improvement. The remaining 52 genes were then processed following these new guidelines. From a practical point of view, one strain was transformed with the 52 plasmids and tested for expression. Only negative targets were tried in the second strain, and so on.

Third Round. A well-known means to exogenously express a recombinant protein in a soluble form is to fuse it to another protein with an intrinsically high solubility.[8,9] This fusion partner is generally removed during the purification procedure by protease digestion, the site of which has been introduced by PCR in the DNA construct.[10] Four different expression plasmids bearing the following tags were available in the laboratory: NusA, maltose binding protein (MBP), glutathione-S-transferase (GST), and thioredoxin (TRX).[8] In all cases, the tag was preceded by His6 and followed by the TEV protease cleavage sequence (borne by the 5′PCR primer), which respectively allowed affinity purification on Ni column and release of the target protein by TEV digestion. In this round, the "two-tubes reaction" had to be used because every target was subcloned into more than one expression vector (Figure 1, blue arrows).

For reasons mentioned earlier, mainly Tb targets underwent this round. Out of 182 targets, 62 were subcloned into MBP vector, 44 into NusA vector, 48 into Trx vector, and 50 into GST vector. Preliminary results indicated that 40 MBP and 4 NusA constructs gave rise to 10 and 2 soluble fusion proteins, respectively. With regard to ASG, 25 ORFs were expressed as an MBP fusion, which gave rise to 14 soluble proteins.

Fourth Round. Because Tb ORFs expressed mostly unfolded proteins in the form of inclusion bodies (IBs) in *E. coli*, it was worth trying refolding IBs by chemical means.[11−13] IBs were dissolved in 6 M guanidinium chloride and purified. The soluble unfolded proteins were refolded using a dilution method in 96-well plates developed in the laboratory (manuscript in preparation). At present, four out of six proteins have been refolded and submitted to crystallization trials, and two gave crystals.

Possible Forthcoming Rounds. Today, only the above four rounds have been used in experiments. However, the following procedures have been successfully applied to reluctant targets elsewhere in the laboratory, but

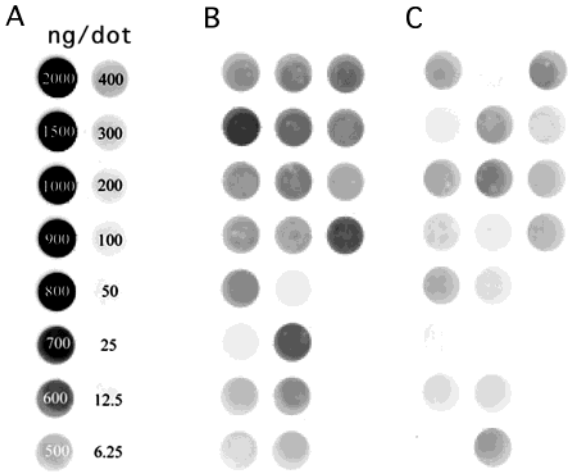

FIGURE 4. Semiquantitative detection of His6-tagged proteins. *E. coli* cells induced to express the protein of interest are lysed, and then the soluble proteins are separated from the unsoluble material by centrifugation and adsorbed on a PVDF membrane using 96-well plates (Millipore ref. MAIPN 0B10) under vacuum. Six His-tagged proteins are detected by incubating blotted proteins with an anti-His antibody coupled to peroxidase (Qiagen) and then with a peroxidase luminescent substrate (ECL, Amersham Biosciences). The resulting luminescent signal is recorded with a computer-driven CCD camera (Kodak). To estimate the amount of recombinant protein, a reference scale made of known amounts of characterized His6-tagged protein is processed in parallel. (A) Luminescent signal produced by the reference scale: the amount of protein, in nanograms per dot, is written on each dot. (B, C) Twenty-four independent *E. coli* clones exhibiting different expression levels: (B) whole lysate and (C) soluble proteins (for a direct comparison, the loading order is the same as in B).

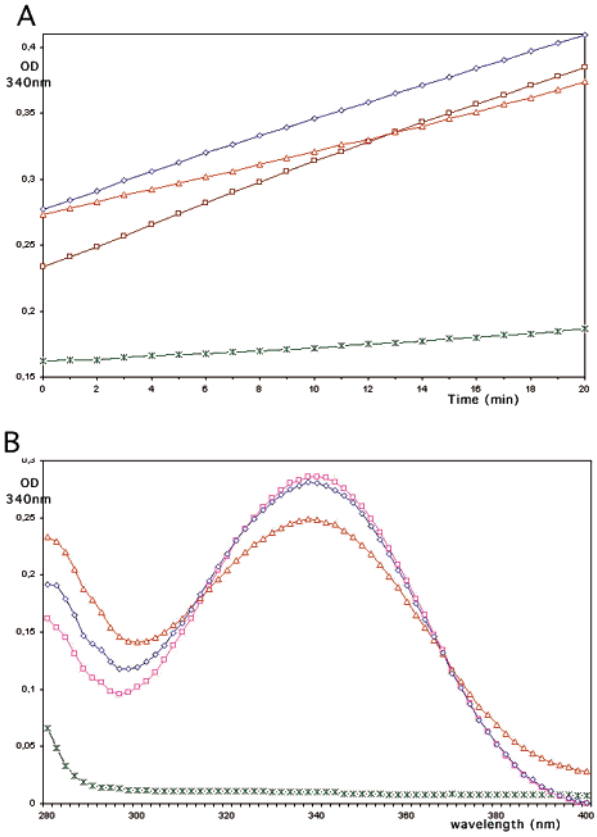

FIGURE 5. Beyond the structures, searching for the molecular function. The sequences of the ORF products were analyzed for the presence of PROSITE "signatures" (http://www.expasy.ch/prosite/) to get information both on possible effectors to be added in the crystallization tests and on the molecular function. In several cases, some "signatures" characteristic of prosthetic group or metal binding were found. Furthermore, out of the first five structures solved, two exhibited active sites which could accommodate NAD or NADP. We have therefore chosen a series of organic molecules (aldehydes, alcohols, sugars, amino acids) suitable as substrates for dehydrogenases; our choice has been driven by the analysis of the specific activity reported for these classes of enzymes in the brenda database (http://www.brenda.uni-koeln.de/). The activity test was set up in 96-well plates. Each well was filled with 100 μL of activity assay mixture containing NAD and a few microliters of enzyme by the TECAN robot. In the above figure, the activity of target 30 with three substrates is depicted. The increase of OD at 340 nm, due to the formation of NADH, was read at 40-s″ interval for 20′ min by a microplate spectrophotometer (Biotek) at 21 °C (A). The final spectrum of each well, including references, was also recorded between 280 and 400 nm, to verify that the increase in OD does actually correspond to the appearance of NADH (B). The specific activity was calculated from the slope of the zero-order kinetics, directly given by the software of the Biotek apparatus, taking into account the extinction coefficients of NADH and the concentration of the enzyme.

not in SG projects, and represent potential fifth and sixth rounds:

(i) In Vitro Expression Using Commercial *E. coli* Extracts (Roche).[14−18] In essence, the latter can express toxic proteins which could account for lack of expression in previous rounds.

(ii) Coexpression with Chaperones. In *E. coli*, endogenous chaperones promote protein folding, and hence protein solubility, during protein synthesis.

Purification of Recombinant Proteins. In view of SAD or MAD experiments at the synchrotron,[19] Se-Met-substituted proteins were produced and purified in similar conditions, using the methionine pathway inhibition method.[20] Since native and Se-Met-substituted proteins were expressed with at least a His6 tag, the mandatory first purification step was passage through a Ni affinity column. The only improvement at this stage was the use of Fast-Flow Chelating Sepharose (Amersham Biosciences), allowing a high flow rate (10 mL/min). Proteins that eluted from the Ni column were further purified by gel filtration on Superdex 200 pg (Amersham Biosciences).

Characterization of Recombinant Proteins. Protein purity and molecular mass were checked by SDS−PAGE and MALDI-TOFF mass spectroscopy,[21] respectively. The mono- or polymeric state of the proteins was determined by dynamic light scattering[22] (DLS) with a DYNAPRO instrument (Protein Solutions). The presence of secondary structures was assessed by circular dichroism[23] with a JASCO 800 spectrometer, and the lack of unfolded domains was confirmed by recording 1D ^{1}H NMR spectra[24] on a Brücker 500DRX spectrometer.

Activity Tests. Activity tests were performed when a putative molecular function could be deduced from the low sequence identity (below 30%) or from the solved structure. The molecular function of several ASG targets

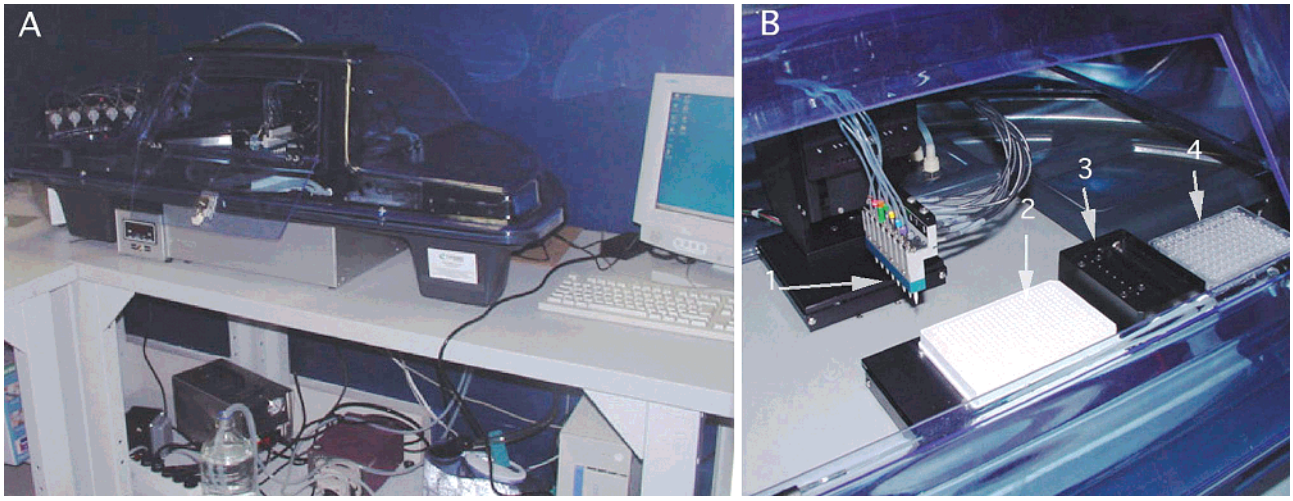

FIGURE 6. The nanodrops dispensing robot (Cartesian Inc.). (A) The carriage and the dispensing tips are contained in a closed box, and the humidity is maintained at 85−90%. (B) The settings used for crystallization: the tips (1) aspirate the protein in the wells of a plate (2) and dispense it in 96 small wells of the Greiner plate (4). The operation can be repeated for three proteins, or for the same protein at three different concentrations. All eight tips are used when the precipitant is aspirated and dispensed. The tips are washed in the "wash station" (3) between pipetting each row of eight wells.

was identified as NAD or NADP redox enzymes. Their specificity was extensively explored (Figure 5).

Crystallization Strategy

Seeking crystallization conditions is a well-known bottleneck in the pipeline of structure production. A detailed description of our crystallization strategy has already been reported.[25] In brief, we had three goals in mind: automate, increase throughput, and reduce protein consumption. These were achieved thanks to three successive generations of robots.

Automate. The first (homemade) robot was able to mimic all the steps normally performed manually, including greasing of the Linbro crystallization plates, dispensing of 2-μL hanging drops on glass coverslips, and the final sealing of the coverslips above the reservoir wells.[25] Although autonomous, this robot did not increase the crystallization throughput because (i) it delivered only a single dispensing at a time, (ii) the Linbro plate accommodated at most 24 samples, and (iii) the whole process was time- (and protein-) consuming.

Increase Throughput. The launch of microplates by Greiner-BioOne with 96 reservoir wells and 288 sitting-drop shelves,[26] devised for the sitting drop vapor diffusion method, resolved these three bottlenecks at once. A TECAN Genesis robot with eight low-volume needles was puchased and used for both loading of the reservoirs and dispensing of 1.5−3-μL sitting drops. Reservoir and protein solutions were mixed together after dispensing to avoid local supersaturation. The crystallization plates were then manually sealed with transparent film and stored at 20 °C. Plates were screened daily by visual inspection. When the first hints became available, improvement of the crystallization conditions was performed by hand by the hanging drop vapor diffusion method in 24-well plates.

Reduce Protein Quantities. SG companies and consortia have reported crystallization in drops in the nanoliter range to be a breakthrough. This aproach both speeds up the kinetics of crystal growth and divides the amount of protein needed for crystallization tests by a factor of at least 10. Toward that goal, we have purchased a dispensing robot (Cartesian Inc.) that uses high-speed microsolenoid valves and therefore is able to dispense "on the fly" drops as small as 10 nL (Figure 6). To prevent the drops from drying out, humidity was maintained at 85−90% within the closed cabinet where the experiments were performed. As the dispensing ceramic tips were calibrated for very low volumes and held at fixed spacing by the dispensing head, the crystallization plate reservoirs had to be filled with the TECAN Genesis robot. The latter transferred the crystallization solutions from 96 15-mL tubes from commercial kits (see below) into the 96 wells of the crystallization plate. The plate was then manually transferred to the Cartesian robot, which performed the following two steps: (i) aspiration of the protein from a remote microplate and dispensing 100−200 nL to the three lateral small wells in each reservoir well, thus allowing for three protein concentrations to be tested in parallel (see Figure 7a), and (ii) aspiration of 100 nL from the reservoir solutions and the dispensing to each of the protein droplets on the lateral shelves.

Another advantage of using "nanodrops" along with high-density crystallization plates was that more screening tests could be assayed. Therefore, our screening experiments now included 408 different conditions. These were obtained by combining five commercially available kits: Structure Screen[27] 1 & 2, Clear Strategy Screen[28] I & II, ZetaSol,[29] Stura Footprint Screen[30] (Molecular Dimensions Ltd., http://www.moleculardimensions.com/), and Wizard screen (Emerald BioStructures, http://www.emeraldbiostructures.com). As mentioned above, each condition was tested at three different protein concentrations, resulting in a total of 1224 crystallization drops, consuming approximatively 1.5 mg of protein.

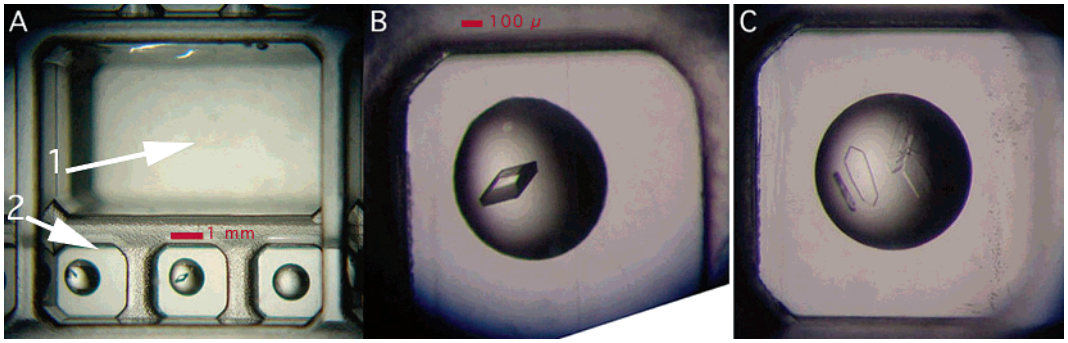

FIGURE 7. Crystallization with 100-nL drops. (A) One of the 96 wells of the Greiner "flat-bottom" crystallization plate. The main well, the reservoir (1), contains the precipitant solution, and the three small lateral wells (2) contain the crystallization drop (here, 100 nL after vapor equilibrium is reached). (B**)** A lysozyme crystal in the crystallization well. Its size (about 0.35 × 0.1 × 0.1 mm) is appropriate for synchrotron data collection. (C) Long, flat crystals of target 87.

Since the "on-the-fly" dispensing mode often led to a bad centering of the drops, we used the slower "step-by-step" dispensing mode. With the "flat-bottom" Greiner plates, the drop was not centered in ~10% of the cases. However, the problem was solved when the "round-bottom" boxes were introduced.

In an initial trial, the nanodrop technology was validated using lysozyme as a positive control.[31] The drops were obtained by mixing 200 nL of a protein solution at half the reported concentration with 100 nL of buffer (Figure 7a). Crystals appeared within 4 h, and 50% of the drops contained nicely formed crystals (Figure 7b). Other crystallization experiments were performed with two other proteins (ASG targets 28 and 87, Figure 7c).

Results of the First and Second Rounds of the Project and Concluding Remarks

As one can see in Figure 3, Gateway recombination cloning was very efficient.

For ASG, the number of expressed ORFs was large, amounting to 94 targets (87%). Sixty-nine proteins (~64% of the targets) were expressed and soluble, which could be considered satisfactory since fusion proteins were not used in rounds 1 and 2. While using several expression strains increased the number of expressed proteins by only 34%, the effect on solubility was more dramatic, with a 100% increase.

The number of proteins brought to crystallization (36) was very inferior to that of soluble proteins identified by screening because parallelization and automation did not apply to protein production. In contrast, the number of crystallized proteins (28) was amazingly high, 78% of the proteins in crystallization test, although only 13 crystals yielded useful data sets: 9 led to structures, the remaining 4 being at the stage of Se-Met protein crystallization. Considering these numbers, the goal of ~20 proteins to be solved should be achieved within 3 years.

Tb program started later than ASG. Out of the 182 ORFs, 85% were cloned and 20% expressed. Most of the proteins were insoluble when only His-tagged. However, four proteins from the fourth screening round reached the crystallogenesis step, and two produced crystals. No diffraction data have been collected yet.

Preparative protein expression and purification were identified as bottlenecks because they were not amenable to parallelization or automation. These could be improved by growing cells in multimicrofermentors[5] (also P. Alzari and J. Bellalou, Pasteur Institute, Paris, personal communication) and purifying proteins with commercial kits such as the 3D kit (Amersham Bioscience). Finally, we have seen that nanodrop technology required about 10 times less protein than classical microdrops. This makes it possible to use expensive in vitro expression (Roche), which could become, in these conditions, an affordable alternative to in vivo expression.

In conclusion, we believe that the iterative strategy of screening rounds of increasing complexity fits well within the scope of medium-throughput SG. In contrast, it would be unrealistic to try to apply it to large-scale projects for practical reasons. As indicated in the Introduction, our next two SG projects concern membrane proteins (MepNet) and viral enzymes (SPINE). It is likely that another combination of screening rounds will be used in either case. We also anticipate that IB refolding (round 4) will be useful in the case of MepNet and eukaryotic expression systems will be required for SPINE.

Avidis SA is greatly acknowledged for making C41/C43 strains available. George Martin (Roche, Berkeley, CA) is acknowledged for the dot-blot by filtration. Arie Geerlof and Günther Stier (EMBL, Heidelberg, Germany) are acknowledged for the NusA, GST, and Trx Gateway vectors. David S. Waugh (MCL, National Cancer Institute, Frederick, MD) is acknowledged for the MBP Gateway vector. Dr. Christopher Barry is thanked for critical reading of the manuscript. This work is part of the ASG program sponsored by the "Ministère de l'Industrie", in collaboration with the IGS laboratory (Marseille) and Aventis.

References

(1) Taylor, W. R. A 'periodic table' for protein structures. *Nature* **2002**, *416*, 657−660.
(2) Erlandsen, H.; Abola, E. E.; Stevens, R. C. Combining structural genomics and enzymology: completing the picture in metabolic pathways and enzyme active sites. *Curr. Opin. Struct. Biol.* **2000**, *10*, 719−730.
(3) Thornton, J. Structural genomics takes off. *Trends Biochem. Sci* **2001**, *26*, 88−90.
(4) Lesley, S. A.; Kuhn, P.; Godzik, A.; Deacon, A. M.; Mathews, I.; Kreusch, A.; Spraggon, G.; Klock, H. E.; McMullan, D.; Shin, T.; Vincent, J.; Robb, A.; Brinen, L. S.; Miller, M. D.; McPhillips, T. M.; Miller, M. A.; Scheibe, D.; Canaves, J. M.; Guda, C.; Jarosze-

wski, L.; Selby, T. L.; Elsliger, M. A.; Wooley, J.; Taylor, S. S.; Hodgson, K. O.; Wilson, I. A.; Schultz, P. G.; Stevens, R. C. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11664−11669.
(5) Lesley, S. A. High throughput proteomics: protein expression and purification in the postgenomic world. *Protein Exp. Purif.* **2001**, *22*, 159−164.
(6) 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.
(7) 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*, 289−298.
(8) Hammarstrom, M.; Hellgren, N.; van Den Berg, S.; Berglund, H.; Hard, T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **2002**, *11*, 313−321.
(9) 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*, 2654−2659.
(10) Dougherty, W. G.;, Carrington, J. C.; Cary, S. M.; Parks, T. D. Mutational analysis of tobacco etch virus polyprotein processing: cis and trans proteolytic activities of polyproteins containing the 49-kilodalton proteinase. *EMBO J.* **1988**, *62*, 2213−2220.
(11) Lindwall, G.; Chau, M.; Gardner, S. R.; Kohlstaedt, L. A. A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng.* **2000**, *13*, 67−71.
(12) Lilie, H.; Schwarz, E.; Rudolph, R. Advances in refolding of proteins produced in E. coli. *Curr. Opin. Biotechnol.* **1998**, *9*, 497−501.
(13) Misawa, S.; Kumagai, I. Refolding of therapeutic proteins produced in Escherichia coli as inclusion bodies. *Biopolymers* **1999**, *51*, 297−307.
(14) Alimov, A. P.; Khmelnitsky, A. Y.; Simonenko, P. N.; Spirin, A. S.; Chetverin, A. B. Cell-free synthesis and affinity isolation of proteins on a nanomole scale. *Biotechniques* **2000**, *28*, 338−344.
(15) Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **1999**, *442*, 15−19.
(16) Yabuki, T.; Kigawa, T.; Dohmae, N.; Takio, K.; Terada, T.; Ito, Y.; Laue, E. D.; Cooper, J. A.; Kainosho, M.; Yokoyama, S. Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J. Biomol. NMR* **1998**, *11*, 295−306.
(17) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* **2002**, *20*, 177−182.
(18) Vincentelli, R.; Abergel, C.; Deregnecourt, C.; Claverie, J.-M.; Monchois, V. Proficient target selection in structural genomics by *in vitro* protein expression on Gateway recombination plasmids. In *Cell-Free Translation Systems*; Spirin, A. S., Ed.; Springer: Berlin, 2002; pp 197−202.
(19) Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991**, *254*, 51−58.
(20) Doublie, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **1997**, *276*, 523−530.
(21) Leushner, J. MALDI TOF mass spectrometry: an emerging platform for genomics and diagnostics. *Expert Rev. Mol. Diagn.* **2001**, *1*, 11−18.
(22) Bernstein, B. E.; Michels, P. A.; Kim, H.; Petra, P. H.; Hol, W. G. The importance of dynamic light scattering in obtaining multiple crystal forms of *Trypanosoma brucei* PGK. *Protein Sci.* **1998**, *7*, 504−507.
(23) van Mierlo, C. P.; Steensma, E. Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *J. Biotechnol.* **2000**, *79*, 281−298.
(24) Montelione, G. T.; Zheng, D.; Huang, Y. J.; Gunsalus, K. C.; Szyperski, T. Protein NMR spectroscopy in structural genomics. *Nat. Struct. Biol.* **2000**, *7* (Suppl.), 982−985.
(25) Sulzenbacher, G.; Gruez, A.; Roig-Zamboni, V.; Spinelli, S.; Valencia, C.; Pagot, F.; Vincentelli, R.; Bignon, C.; Salomoni, A.; Grisel, S.; Maurin, D.; Huyghe, C.; Johansson, K.; Grassick, A.; Roussel, A.; Bourne, Y.; Perrier, S.; Miallau, L.; Cantau, P.; Blanc, E.; Genevois, M.; Grossi, A.; Zenatti, A.; Campanacci, V.; Cambillau, C. A Medium Throughput Crystallization Approach. *Acta Crystallogr.* **2000**, *D58*, 2109−2115.
(26) Mueller, U.; Nyarsik, L.; Horn, M.; Rauth, H.; Przewieslik, T.; Saenger, W.; Lehrach, H.; Eickhoff, H. J. Development of a technology for automation and miniaturization of protein crystallization. *Biotechnology* **2001** *85*, 7−14.
(27) Jancarik, J.; Kim, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **1991**, *24*, 409−411.
(28) Dauter, Z.; Dauter, M.; Rajashankar, K. R. *Acta Crystallogr.* **2000**, *D56*, 232−237.
(29) Riès-Kautt, M.; Ducruix, A. Inferences from physico-chemical studies of crystallogenesis and the precrystalline stage. *Methods Enzymol.* **1997**, *276*, 23−59.
(30) Stura, E. A.; Nemerow, G. R.; Wilson, I. A. Strategies in the crystallization of glycoproteiens and protein complexes. *J. Crystal Growth* **1992**, *122*, 273−285.
(31) Stevens, C. O.; Bergstrom, G. R. The multiple nature of crystalline egg-white lysozyme. *Proc. Soc. Exp. Biol. Med.* **1967**, *124*, 187−191.

AR010130S