

Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE

Pedro M. Alzari,^a H. Berglund,^b
N. S. Berrow,^c E. Blagova,^d
D. Busso,^e C. Cambillau,^f
V. Campanacci,^f E. Christodoulou,^g
S. Eiler,^e M. J. Fogg,^d G. Folkers,^h
A. Geerlof,ⁱ D. Hart,^j A. Hauz,^a
M. D. Herman,^b S. Macieira,^k
P. Nordlund,^l A. Perrakis,^g
S. Quevillon-Cheruel,^m
F. Tarandeu,^j H. van Tilbeurgh,^m
T. Unger,ⁿ M. P. A. Luna-Vargas,^g
M. Velarde,^k M. Willmannsⁱ and
Raymond J. Owens^{c*}

^aUnité de Biochimie Structurale, Institut Pasteur, 25–28 Rue du Dr Roux, 75724 Paris CEDEX 15, France,

^bDepartment of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-109 51 Stockholm, Sweden,

^cOxford Protein Production Facility and Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, England,

^dYork Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England, ^eInstitut de Génétique et de Biologie

Moléculaire et Cellulaire, 1 Rue Laurent Fries, BP 163, 67404 Illkirch CEDEX, France, ^fArchitecture et Fonction des Macromolécules Biologiques UMR6098, CNRS/Universités de Provence/Université de la Méditerranée

Parc Scientifique et Technologique de Luminy, Case 932163, Avenue de Luminy, 13288 Marseille

CEDEX 09, France, ^gDivision of Molecular

Carcinogenesis, Netherlands Cancer Institute,

Plesmanlaan 121, 1066 CX Amsterdam,

The Netherlands, ^hBijvoet Center for Biomolecular

Research, NMR Spectroscopy, Utrecht University,

Padualaan 8, 3584 CH Utrecht, The Netherlands,

ⁱEMBL Hamburg Outstation, Notkestrasse 85,

D-22603 Hamburg, Germany, ^jEMBL Grenoble, c/o ILL,

BP 181, 6 Rue Jules Horowitz, F-38042 Grenoble

CEDEX 9, France, ^kMax-Planck-Institute of Biochemistry,

Department of Proteomics and Signal Transduction,

Am Klopferspitz 18, 82152 Martinsried, Germany,

^lDepartment of Biotechnology, Royal Institute of

Technology, AlbaNova University Centre,

S-10691 Stockholm, Sweden, ^mInstitut de Biochimie et

de Biophysique Moléculaire et Cellulaire, UMR8619,

Bâtiment 430, Université de Paris-Sud, 91405 Orsay

CEDEX, France, and ⁿThe Israel Structural Proteomics

Centre, The Department of Structural Biology,

Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: ray@strubi.ox.ac.uk

The implementation of high-throughput (HTP) cloning and expression screening in *Escherichia coli* by 14 laboratories in the Structural Proteomics In Europe (SPINE) consortium is described. Cloning efficiencies of greater than 80% have been achieved for the three non-ligation-based cloning techniques used, namely Gateway, ligation-independent cloning of PCR products (LIC-PCR) and In-Fusion, with LIC-PCR emerging as the most cost-effective. On average, two constructs have been made for each of the approximately 1700 protein targets selected by SPINE for protein production. Overall, HTP expression screening in *E. coli* has yielded 32% soluble constructs, with at least one for 70% of the targets. In addition to the implementation of HTP cloning and expression screening, the development of two novel technologies is described, namely library-based screening for soluble constructs and parallel small-scale high-density fermentation.

1. Introduction

High-throughput (HTP) sequencing of eukaryotic, viral and bacterial genomes is providing a huge database of proteins with potential for structure–function analysis. In response to this opportunity, structural proteomics projects have been initiated worldwide with the aim of establishing HTP structure determination on a genome-wide scale. Crucial to this effort has been the development of production technologies for the HTP cloning, expression and purification of recombinant proteins. In all projects, there has been an emphasis on parallel processing for molecular cloning, expression and purification. This has been driven by the need to accommodate relatively large numbers of potential targets for structural biology at an acceptable cost and has led to varying degrees of automation. Most of the groups involved have set up semi-automated liquid-handling systems to carry out some or all of their protocols. However, the protocols can be carried out equally well by hand with appropriate equipment, e.g. multi-channel pipette dispensers. The motivation to implement automation is largely to enable processes to be scaleable and run routinely as error-free operations, leading to greater reproducibility compared with procedures carried out manually. The EU-funded Structural Proteomics In Europe (SPINE; <http://www.spineurope.org>) programme has provided the opportunity for method developments at a number of European centres and the exchange of experience during the 3 y of the project.

For structural biology, expression of recombinant proteins in *Escherichia coli* remains the most widely used approach for obvious reasons of speed, ease of use and low cost. The SPINE project has been no exception and approximately 200 new structures have been obtained from proteins produced in

Received 17 November 2005

Accepted 31 July 2006

Table 1

Summary of the technology developments of the contributing SPINE laboratories.

SPINE site	Authors	Cloning methods	Screening methods	Liquid-handling system(s)	Other methods	Applications in SPINE	References
Amsterdam	E. Christodoulou, M. P. A. Luna-Vargas, A. Perrakis	LIC-PCR	Ni-NTA magnetic beads, filtration/gels	Tecan	In-house vectors	Human	
Grenoble	F. Tarendeau, D. Hart	RE-based	Colony blot		Library method	Human, viral	Hart & Tarendeau (2006)
Hamburg	A. Geerlof, M. Wilmanns	Gateway, RE-based	Ni-NTA resin/gels	N/A	New Gateway vector suite novel strains	Human	
Marseille	V. Campanacci, C. Cambillau	Gateway	Filter dot-blot/gels	Tecan	Refolding screen	Bacterial, viral	Vincentelli <i>et al.</i> (2003, 2004, 2005)
Munich	S. Macieira, M. Velarde	Gateway	Ni-NTA resin/GST-resin/gels	N/A	New Gateway vector suite, N-terminal sequencing	Human	
Orsay	S. Cheruel, H. van Tilbeurgh	RE-based	Ni-NTA resin/gels	MWG	Chaperone co-expression/refolding screen/library methods	Yeast	Quevillon-Cheruel <i>et al.</i> (2004), Teplyakov <i>et al.</i> (2003)
Oxford	N. Berrow, R. Owens	Gateway, In-Fusion	Ni-NTA magnetic bead/gels	Qiagen 8000 MWG Theonyx	New In-Fusion vector suite	Human, viral, bacterial	In preparation
Paris	P. Alzari, A. Haouz	N/A	Gels	N/A	Parallel fermenters, cell-free expression for construct optimization	<i>M. tuberculosis</i>	Betton (2004, in preparation)
Rehovot	T. Unger	RE-based	Ni-NTA resin/gels	N/A		Human	
Stockholm	P. Nordlund, M.-D. Herman, H. Berglund	Gateway	Filter/dot-blot/gels	Qiagen 8000	Parallel fermenters	Human, bacterial	Knaust & Nordlund (2001), Hammarstrom <i>et al.</i> (2002), Cornvik <i>et al.</i> (2005)
Strasbourg	D. Busso, S. Eiler	Gateway	Ni-NTA resin/gels	Tecan	New GW vector suite	Human	Busso, Delagoutte-Busso <i>et al.</i> (2005), Busso, Poussin-Courmontagne <i>et al.</i> (2005)
Utrecht	G. Folkers	LIC-PCR enzyme-free	Ni-NTA magnetic bead/gels	Hamilton Star	2D NMR screening	Human	Folkers <i>et al.</i> (2004)
York	M. Fogg, E. Blagova	LIC-PCR	Ni-NTA resin/gels	N/A	In-house vector	Bacterial	

E. coli. The development and implementation of HTP approaches to cloning and expression in *E. coli* have made a major contribution to this output. In this article, we review the technical developments that have come from SPINE and provide protocols for carrying out cloning and expression screening, including refolding, in a relatively HTP and parallel approach. Apart from sponsoring the establishment of Europe-wide HTP activity, the SPINE programme has enabled the development of novel technology focused at specific sites. In the context of protein production in *E. coli*, two examples will be reported which demonstrate such innovation. Firstly, a library-based approach to high-density screening for the expression of soluble fragments has been set up at the EMBL laboratories in Grenoble. This builds on to recent developments in which random screening strategies are being used to define solubly expressing constructs (reviewed in Hart & Tarendeau, 2006). Secondly, novel systems for small-scale parallel high-density fermentation of *E. coli* have been developed at the Pasteur Institute (Bellalou *et al.*, manuscript in preparation; Hedrén *et al.*, 2006).

2. Experimental

2.1. Vector construction

Amongst SPINE groups, ligation-dependent, ligation-independent and site-specific recombinatorial cloning have

been used to construct the expression vectors required for protein production (see Table 1). Of these methods, the Gateway system of recombinatorial cloning has been the most widely used (six laboratories). This is a modification of the recombination system of phage λ (Hartley *et al.*, 2000; Walhout *et al.*, 2000) and the Gateway system utilizes a minimum set of components of the λ system for *in vitro* transfer of DNA. Directional cloning of the DNA insert is ensured by using two nearly identical but non-compatible versions of the λ att recombination site. The system works by a two-step process: firstly in the BP reaction PCR products are cloned into a generic 'entry vector', whilst in the second (LR) step a recombination reaction transfers the gene of interest into the chosen expression vector. Implementation of Gateway in SPINE has been according to the manufacturer's protocol (<http://www.invitrogen.com>) with reaction volumes generally down-scaled, e.g. to 5 μ l (BP) and 10 μ l (LR) (Vincentelli *et al.*, 2003; Busso *et al.*, 2005). In all cases, new and improved Gateway-compatible vectors have been constructed to customize the system (Table 2). An example of the use the Gateway system in HTP mode comes from the OPPF, in which 342 PCR products have been cloned into the Gateway expression vector pDEST 14 (Invitrogen) with a success rate of 86% as assessed by PCR screening of constructs. This is consistent with reports from other groups outside SPINE (e.g. Thao *et al.*, 2004). It appears that the overall cloning efficiency

Table 2

List of *E. coli* expression vectors constructed by SPINE laboratories.

Vector sequences are available from the authors and have been deposited in GenBank.

Vector name	Description	Originator
pDESTN-His15	Modified pDEST15 (Invitrogen) to incorporate N-His ₆ upstream of GST	Oxford
pET-10AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-His ₆ and C-His ₆ tags	Hamburg
pETG-20AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-thioredoxin-His ₆ and C-His ₆ tags	Hamburg
pETG-30AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-His ₆ -GST and C-His ₆ tags	Hamburg
pETG-40AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-MBP and C-His ₆ tags	Hamburg
pETG-41AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-His ₆ -MBP and C-His ₆ tags	Hamburg
pETG-50AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-DsbA-His ₆ and C-His ₆ tags	Hamburg
pETG-52AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-leaderless-DsbA-His ₆ and C-His ₆ tags	Hamburg
pETG-60AEMBL	pET-22b(+) (Novagen) adapted for Gateway incorporates N-NusA-His ₆ and C-His ₆ tags	Hamburg
pTH10	Gateway adapted from pT7-ZZA, N-terminal Z-domain fusion, PreScission (3C) cleavage site	Stockholm
pTH18	Gateway adapted from pET21, N-terminal GB1 fusion, PreScission cleavage site	Stockholm
pTH19	Gateway adapted from pET 15b, N-terminal His ₆ , thrombin cleavage site	Stockholm
pTH1	Gateway adapted from pMAL-c, N-terminal MBP fusion, factor Xa cleavage site	Stockholm
pTH2	Gateway adapted from pET-43a, N-terminal NusA fusion	Stockholm
pTH3	Gateway adapted from pGb1, N-terminal GB1 fusion	Stockholm
pTH5	Gateway adapted from pT7-ZZA, N-terminal ZZ-domain fusion, Genenase 1 cleavage site	Stockholm
pTH6	Adapted from pDEST15 GST PreScission cleavage site	Stockholm
pTH7	Gateway adapted from pET43a, N-terminal NusA fusion, PreScission cleavage site	Stockholm
pTH8	Adapted from pDEST16, N-terminal thioredoxin fusion, PreScission cleavage site	Stockholm
pTH24	Adapted from pET-DEST42, C-terminal His ₆ tag	Stockholm
pTH27	Gateway adapted from pET-21, N-terminal His ₆ tag	Stockholm
pTH28	Gateway adapted from pET-21, N-terminal thioredoxin fusion, PreScission cleavage site	Stockholm
pTH29	Gateway adapted from pET-21, N-terminal GST, PreScission cleavage site, His ₆ tag	Stockholm
pTH30	Gateway adapted from pET-21, N-terminal Z-domain PreScission cleavage site, His ₆ tag	Stockholm
pTH31	Gateway adapted from pET11c, C-terminal EGFP fusion	Stockholm
pTH34	Gateway adapted from pET-21, N-terminal GB1 domain, PreScission cleavage site, His ₆ tag	Stockholm
pTH35	Gateway adapted from pET-21, N-terminal GST fusion, PreScission cleavage site	Stockholm
pTH36	Gateway adapted from pET-21, N-terminal thioredoxin fusion, PreScission cleavage site	Stockholm
pTH38	Gateway adapted from pET-43a, N-terminal NusA, PreScission cleavage site, His ₆ tag	Stockholm
pHGWA	pET-22b adapted for Gateway incorporates N-His ₆ and C-His ₆ tags	Strasbourg
pHMGGWA	pET-22b adapted for Gateway incorporates N-His ₆ -GST and C-His ₆ tags	Strasbourg
pHMGWA	pET-22b adapted for Gateway incorporates N-His ₆ -MBP and C-His ₆ tags	Strasbourg
pHNGWA	pET-22b adapted for Gateway incorporates N-His ₆ -NusA and C-His ₆ tags	Strasbourg
pHXGWA	pET-22b adapted for Gateway incorporates N-His ₆ thioredoxin and C-His ₆ tags	Strasbourg
p0GWA	pET-22b adapted for Gateway incorporates C-His ₆ tag	Strasbourg
p0GGWA	pET-22b adapted for Gateway incorporates N-GST and C-His ₆ tags	Strasbourg
p0MGWA	pET-22b adapted for Gateway incorporates N-MBP and C-His ₆ tags	Strasbourg
p0NGWA	pET-22b adapted for Gateway incorporates N-NusA and C-His ₆ tags	Strasbourg
p0XGWA	Modified pET-22b for Gateway incorporates N-thioredoxin and C-His ₆ tags	Strasbourg
pG4_casB	Modified pGEX-4T-2 (Pharmacia) for Gateway Tac promoter N-GST-thrombin cleavage site	Munich
pG5_casA	Modified pGEX-5X-3(Pharmacia) for Gateway Tac promoter N-GST factor Xa cleavage site	Munich
pI7_casB	Modified pASK-IBA7 for Gateway (IBA Institute) tet promoter N-Strep-TagII and factor Xa cleavage site	Munich
pTYB2_casC	Modified pTYB2 (New England BioLabs) for Gateway T7/lac promoter C-intein self-cleaving tag	Munich
pTrcHisA_casB	Modified pTrcHisA (Invitrogen) for Gateway Trc promoter N-His ₆ enterokinase cleavage site	Munich
pET-46NKI/LIC	Modified pET-46Ek/LIC vector (Novagen) incorporating 600 bp insert, zero background, N-His ₆ tag and enterokinase cleavage site, Amp ^R or Kan ^R	Amsterdam
pET-22NKI/LIC (construction in progress)	Modified pET-22b vector (Novagen) incorporating 600 bp insert, zero background, no N-terminal tag, choice of no or C-terminal His ₆ tag, Amp ^R or Kan ^R	Amsterdam
pET-28NKI/LIC (construction in progress)	Modified pET-28a vector (Novagen) incorporating 600 bp insert, zero background, N-His ₆ tag and HRV 3C cleavage site, Amp ^R or Kan ^R	Amsterdam
pET-YSBLIC	pET-28a adapted for LIC incorporates N-His ₆ tag	York
pET-YSBLIC3C	pET-28a adapted for LIC incorporates N-His ₆ tag and 3C protease cleavage site	York
pOPINA	pET-28a modified for In-Fusion incorporates either N-His ₆ or C-His tags depending upon site of cloning	Oxford
pOPINB	pET-28a modified for In-Fusion includes N-His ₆ tag and 3C protease cleavage site or C-His tags depending upon site of cloning	Oxford

of the Gateway system is largely determined by the BP reaction, since the LR recombination step is generally 100% efficient.

The major alternative to Gateway in SPINE has been the use of ligation-independent cloning of PCR products (LIC-PCR), which was developed over 10 y ago (Aslandis & deJong, 1990; Haun *et al.*, 1992). LIC-PCR is based on the use of T4 DNA polymerase in the presence of a single deoxyribonucleotide to produce 12–15 bp overhangs in a PCR product

that are complementary to sequences generated in the recipient vector. These extensions anneal sufficiently strongly to allow transformation of *E. coli* without the need for a ligation step which is carried out by repair enzymes in the host. LIC-PCR has been successfully implemented in 96-well format by three groups using variants of the original protocol. The LIC-PCR procedure is exemplified by the protocol developed by the York Structural Biology Laboratory (York SPINE partner; Fogg *et al.*, unpublished work) and is representative of those

5' - ATGGGCAGCAGCCATCATCACCACCACCACAGGCGCGCCTTCTCCTCACTGTTCCAGGGGGCCC - 3'
 3' - TACCCGTCGTCGGTAGTAGTAGTAGTGTCCGCGCGGAAGAGGAGTGACAAGGTCCCCGGG - 5'
 MetGlySerSerHisHisHisHisHisHisHis ↑ BseRI
 Non-cleavable His tag LIC site

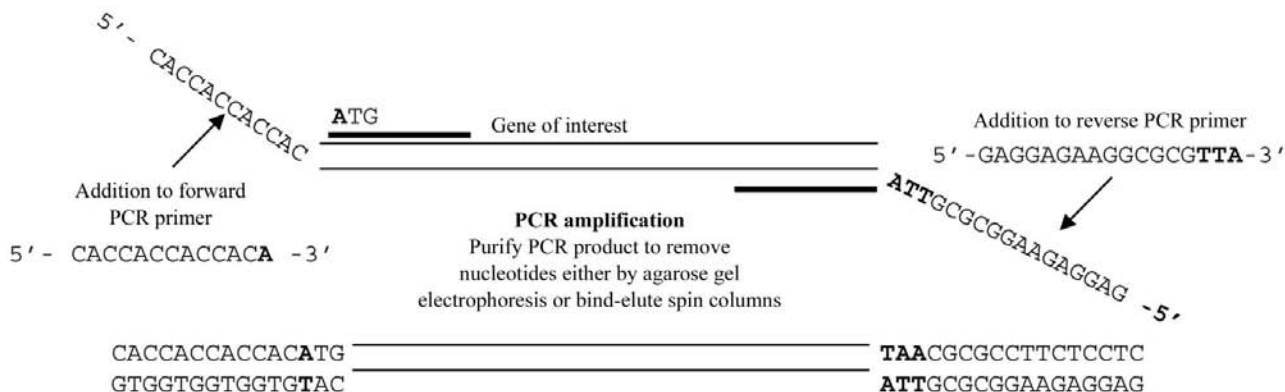
BseRI digestion
 Digest 50 µg vector linearized with 200 U BseRI and purify by agarose gel electrophoresis

5' - GCAGCAGCCATCATCACCACCACCACAGG - 3' 5' - CGCGCCTTCTCCTCACTGTTCCAGGGGGCCC - 3'
 3' - CGTCGTCGTCGGTAGTAGTGGTGGTGGTGT - 5' 3' - CCGCGCGGAAGAGGAGTGACAAGGTCCCCGGG - 5'

T4 polymerase/dTTP treatment
 Incubate 4.0 pmol linearized vector with 20 U T4 polymerase in 1× T4 polymerase buffer containing 2.5 mM dTTP and 5 mM DTT in a total volume of 0.4 ml for 30 min at 295 K followed by 20 min at 348 K. Purify by bind-elute spin column and store at 253 K.

GCAGCAGCAGCCATCAT - 3' 5' - CGCGCCTTCTCCTCACTGTTCCAGGGGGCCC
 CGTCGTCGTCGGTAGTAGTGGTGGTGGTGT - 5' 3' - TGACAAGGTCCCCGGG

(a)



T4 polymerase/dATP treatment
 Incubate 0.2 pmol PCR product with 1 U T4 polymerase in 1× T4 polymerase buffer containing 2.5 mM dTTP and 5 mM DTT in a total volume of 0.02 ml for 30 min at 295 K followed by 20 min at 348 K

CACCACCACCAC**ATG** TAA
 AC ATTGCGCGGAAGAGGAG

Annealing vector and insert
 Anneal 0.02 pmol T4 polymerase-treated insert with 50 ng T4-treated vector for 5 min at 295 K. Add EDTA to a final concentration of 6 mM, incubate for 5 min at 295 K and transform cells with 1-2 µl of the annealing reaction

GCAGCAGCAGCCATCAT - 3' 5' - CGCGCCTTCTCCTCACTGTTCCAGGGGGCCC
 CGTCGTCGTCGGTAGTAGTGGTGGTGGTGT - 5' 3' - TGACAAGGTCCCCGGG

(b)

Figure 1

Schematic representation of the LIC-PCR protocol. The procedure for cloning into the LIC-PCR vector pET-YSB LIC (Table 2) is outlined. In (a) the method for preparing the vector is described and in (b) the preparation of the PCR fragment and annealing to the vector is shown.

used by other groups in SPINE (Fig. 1). The York LIC-PCR vector has been used to clone 263 PCR products with a success rate of 85%. Finally, a second method for LIC cloning has been implemented by the Oxford partners during the course of SPINE using the In-Fusion technology (Clontech; Berrow *et al.*, manuscript in preparation). The In-Fusion method is both insert-sequence independent and enables cloning of PCR products directly into any cloning or expression vector *via* primer extensions defined by the user. Integration of the insert and vector DNA is achieved in a single-step reaction catalysed by the proprietary In-Fusion enzyme (<http://www.clontech.com>). In-Fusion cloning has been used by the Oxford partners to clone a set of 347 PCR products with an efficiency of 89% as assessed by PCR screening of recombinant clones. This is consistent with an efficiency of 85% reported by Hartman *et al.* (2005) for the cloning of 874 PCR amplicons by In-Fusion. A generic workflow for HTP cloning is shown in Fig. 2 based on the working practice of the Oxford partners.

The 10–15% cloning failure observed for all the strategies outlined above probably arises from a proportion of poor-quality PCR products in the experiments. Typically, in HTP mode only one PCR condition for reaction volume, cycling parameters, amount of template and primer is used to amplify all inserts. This is chosen to give the best overall coverage, but may compromise the yield of particular target sequences leading to subsequent cloning failures. However, these can frequently be rescued in an individual experiment in which PCR conditions are optimized for specific sequences.

2.2. Expression screening in 96-well plate format

A major feature of the HTP protein-production pipelines developed in SPINE is the inclusion of a screening step on a relatively small scale to identify constructs suitable in terms of soluble protein yield for scale-up and subsequent protein purification. Given the relatively high cost of the latter steps in terms of time and resources, the screening stage is seen as crucial to the overall process (Table 1). All groups make use of T7-based vectors and the BL21 (DE3) strain of *E. coli* and its derivatives for expression screening, although the number of different strains evaluated in parallel varies between groups (Aslandis & deJong, 1990; Haun *et al.*, 1992; Quevillon-Cheruel *et al.*, 2003; Folkers *et al.*, 2004; Busso *et al.*, 2005; Vincentelli *et al.*, 2005). Typically, small-scale expression screens are carried out in 96- or 24-well deep-well plates using enriched complex media *e.g.* TB, 2YT, GS96 (QBiogene) to ensure maximum biomass. Generally, these media support growth to optical densities (OD) of 5–10 OD₆₀₀ units compared with 2–3 OD₆₀₀ units in standard Luria Broth (LB). Expression is induced by either the addition of IPTG (0.1–1.0 mM) or by the autoinduction method of Studier (2005).

The starting point for the expression assay is lysing the cells after harvesting and then separation of soluble from insoluble fractions. Cell lysis is carried out using standard protocols by either a freeze–thaw cycle followed by treatment with DNase/lysozyme or by sonication with/without lysozyme or using

commercial detergent-based lysis reagents, *e.g.* BugBuster (Merck). Chemical methods lend themselves to 96-well formats, although sonicators which can accommodate a 96-well plate are available (Misonix) and used by the Strasbourg group for cell lysis in this format. Generally, soluble products are then purified on a small scale using either IMAC magnetic beads (*e.g.* Qiagen, Novagen) or IMAC resins in filter plates *via* a His tag on the protein. Standard manufacturers' protocols are used for this step, with most groups using robotic liquid-handling systems to automate the process (Table 1). Two main assay formats have been adopted for detecting soluble protein expression, namely immunodetection using dot-blots of either lysates or IMAC-purified soluble proteins (Knaust & Nordlund, 2001; Cornvik *et al.*, 2005; Vincentelli *et al.*, 2005) or conventional SDS–PAGE of samples (Folkers *et al.*, 2004; Busso *et al.*, 2005). A generic workflow for expression screening based on the Oxford procedures is shown in Fig. 2. Details of each protocol are given in the accompanying paper (Berrow *et al.*, 2006), in which a set of 96 expression constructs were screened by different SPINE groups to benchmark the protocols against each other.

Proteins assessed by screening to be suitable for scale-up in terms of quality and quantity, typically >95% pure by SDS–PAGE with a projected yield of $\geq 0.5 \text{ mg l}^{-1}$, were purified from 1–2 l *E. coli* cultures for subsequent structural studies.

2.3. Deletion-library construction and screening

Both the EMBL Grenoble and the Stockholm groups have developed methodologies for the construction and screening of gene-deletion libraries for identification of truncated reading frames that express soluble proteins. The Stockholm group combined the Erase-a-base protocol (Promega Corp.) with a colony-lift procedure to identify such constructs (Cornvik *et al.*, 2005). The Grenoble group have adapted library-handling and screening robotics, originally developed for genomics applications, to the purpose of systematically screening gene-truncation and gene-fragmentation libraries. Both groups analyse protein expression at the colony level, which avoids the logistical difficulties of simultaneous protein expression and testing of solubility for thousands of different clones in multi-well plate format. The Grenoble process, ESPRIT (expression of soluble proteins by random incremental truncation), involves the robotic picking of 30 000 individual constructs into 384-well plates followed by the printing of high-density colony arrays on nitrocellulose filters. Colonies are grown at three separate temperatures and protein expression is induced by shifting the filters onto LB agar containing inducing agent (IPTG or arabinose). After lysis and fixing of cellular proteins onto the nitrocellulose filter, soluble protein expression is detected using a fused linear peptide that is efficiently post-translationally modified *in vivo* only if the protein is both soluble and stable (manuscript in preparation). The geometric format of the printed colony filters permits the use of array-analysis software, originally designed for DNA arrays, for quantification of

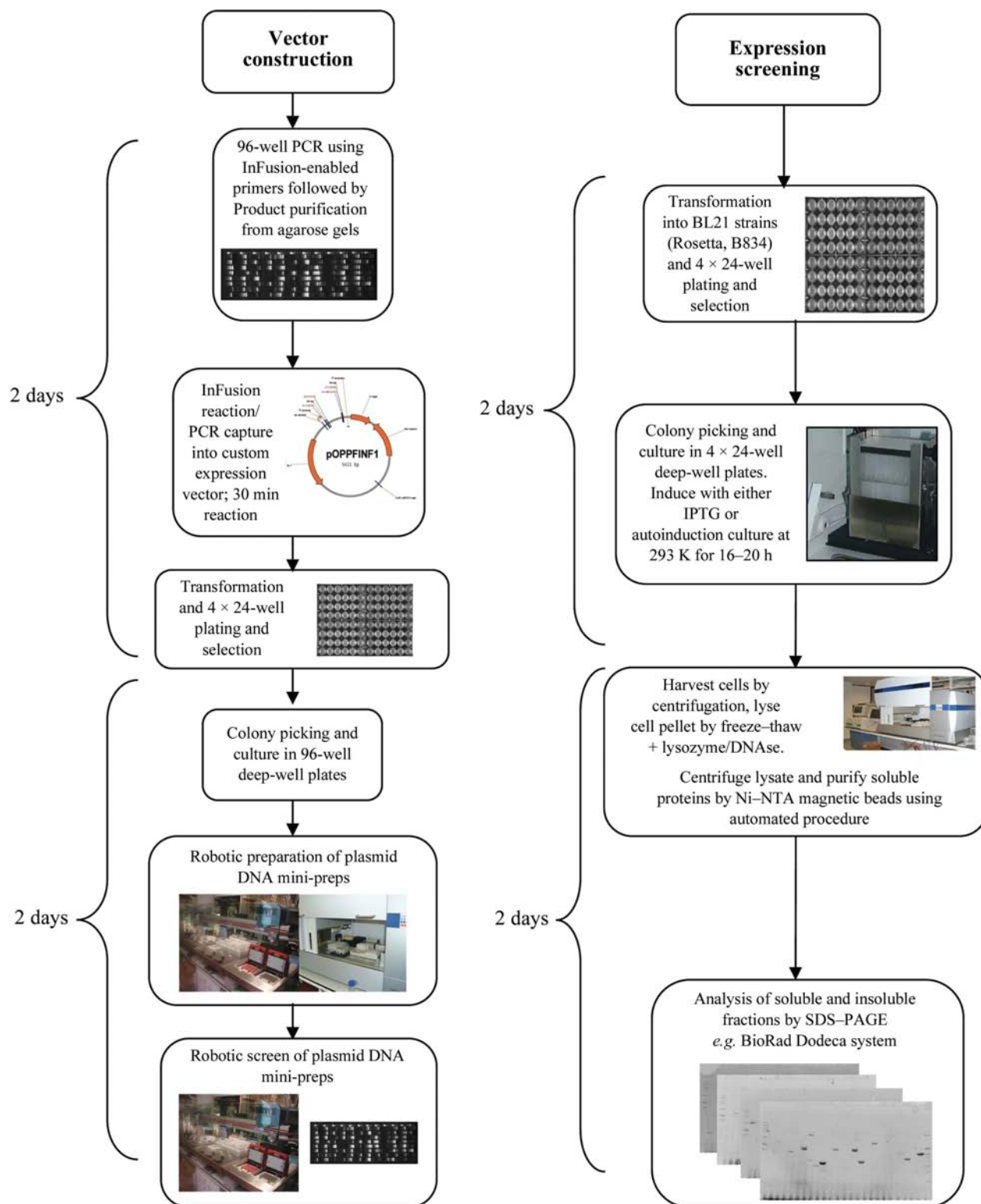


Figure 2

Generic workflow for HTP cloning and expression screening. The scheme is based on the current procedures of the Oxford Protein Production Facility and shows that both cloning and expression screening can be carried out in two working weeks with certain steps automated. PCR reactions for both amplifying the target-gene sequences and for screening mini-preps following cloning (robotic screen of plasmid DNA mini-preps) are carried out in a semi-automated procedure using a PCR cycler integrated into a liquid- and microplate-handling system (MWG Theonyx). Small-scale plasmid DNA mini-preps are prepared automatically using the Qiagen 8000 instrument and associated protocol/reagents. Expression vectors are verified by PCR screening and sequencing on one strand only from a single primer (T7 forward). Transformation of *E. coli* strains both for cloning and expression screening is carried out manually with cells plated out on standard 24-well plates (15 mm diameter well) to give a density of approximately 10–20 colonies per well determined empirically. Colony picking and culture in deep-well blocks is also carried out by hand. Expression screening is semi-automated with small-scale Ni–NTA purification of soluble proteins carried out using the Qiagen 8000 instrument and associated protocol/reagents.

signal intensity and ranking of clones. Positive clones are then robotically re-arrayed into a single 96-well plate for further confirmation of protein solubility and characterization of the domain boundaries by DNA sequencing.

2.4. Refolding from inclusion bodies

A number of SPINE partners have developed protocols to recover protein from inclusion bodies. Some of the approaches taken by the Marseille, Orsay and Weizmann groups have been described in detail elsewhere (Vincentelli *et al.*, 2004; Tresaugues *et al.*, 2004; Albeck *et al.*, 2005). Briefly, for those proteins that are expressed exclusively in the insoluble fraction based on the screen, washed inclusion bodies are prepared and solubilized in guanidinium hydrochloride. The His-tagged proteins are then purified under denaturing conditions and protein refolding attempted by dilution into a set of refolding buffers in a 96-well format. Folding of protein is followed by measuring light scattering at 350 or 390 nm and the screen has been fully automated (Vincentelli *et al.*, 2004). Further assessment of refolding and validation of any hits requires scale-up of the process under the conditions identified by the screen and biophysical characterization of the products. Generic methods used to assess authentic folding include size-exclusion chromatography (SEC; monodispersity), circular dichroism (secondary structure) and dynamic light scattering (DLS; monodispersity). Similarly at the Weizmann, His-tagged protein is partially purified in the denatured state by capture on nickel-NTA before dilution

into various buffers containing additives such as salts, polar additives, osmolites, detergent additives and chaotropes at three different pH values (Albeck *et al.*, 2005).

As an alternative to screening for refolding, a high level of success for particular groups/classes of proteins is possible using a single condition. For example, a generic approach has been used by the Oxford group for extracellular proteins stabilized by disulfide bonds (Gao *et al.*, 1998). The procedure involves preparation of inclusion bodies on a large scale (from 2 l cultures) followed by solubilization in 6 M guanidine-HCl, 50 mM Tris pH 8, 100 mM NaCl, 10 mM EDTA and 10 mM DTT. Denatured protein is then rapidly diluted into a large volume of refolding buffer containing 200 mM Tris-HCl pH 8, 10 mM EDTA, 1 M L-arginine, 0.1 mM PMSF, 6.5 mM cysteamine and 3.7 mM cystamine.

2.5. High-density parallel fermentation

Most groups have a target yield of at least 5 mg purified protein per litre of *E. coli* culture and with the general take-up of nanolitre crystallization methods this is usually sufficient protein for crystallization screening (see Berry *et al.*, 2006). Thus, for many projects growing cells in simple shake-flask cultures to 1 l is usually adequate and several different production runs can easily be handled in parallel. However, for some lower yielding but high-value proteins larger volumes would be required and the use of fermenter systems in which growth and productivity can be controlled and optimized becomes appropriate. In the SPINE consortium, dedicated faculties have been established at the EMBL, Hamburg to carry out high-density fermentations ($OD_{600} > 100$) up to the 1 l scale using four individual fermenters. In addition, the Strasbourg group use a commercially available system comprising a battery of six small (0.5 l working volume) fermenters (SixFors, Infors). However, there remains the need for a more highly parallel system for high-density fermentation. With these requirements in mind, the Pasteur group have built a computer-controlled battery comprising eight miniaturized fermenters of 80 ml capacity. The reactors consist of glass vessels with a square section, which enables continuous monitoring of the optical density of the cultures by means of an external mobile optical sensor which moves from one reactor to the next. The temperature of each reactor is controlled by an internal probe and by Peltier elements, which can be programmed independently. The fermenters can be fitted with pH and pO_2 probes and with an automated injection system for adding inducer at a defined stage of the culture. Highly efficient oxygen-enriched aeration provided by sintered glass spargers enables the batch-wise cultivation of *E. coli* to high cell densities using a home-developed enriched medium. OD_{600} values of 60–100 are routinely achieved. Finally, a second parallel fermentation system has been developed by the Stockholm group which is similar to the SixFors but comprises 12 culture vessels (see Hedrén *et al.*, 2006).

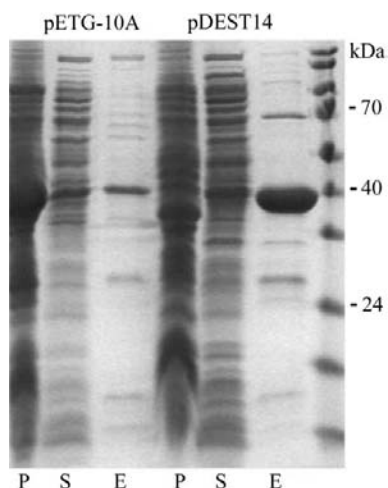


Figure 3

Expression of the protein W00005 in either pETG-10A (*5'* *att* recombinatorial site included in the translation product) or pDEST14 (*5'* *att* recombinatorial site excluded from the translation product). A parallel expression experiment was performed in *E. coli* BL21(DE3) using the two Gateway-compatible vectors, pETG-10A (Table 2) and pDEST14 (InVitrogen). The bacterial cultures were grown at 310 K to OD of 0.6 at 600 nm and induced with 50 μ M IPTG at 303 K for 4 h. Equal amount of cells (based on the OD at 600 nm) were withdrawn for solubility analysis. Cells were lysed by sonication and soluble (lane S) and insoluble fractions (lane P) were separated by centrifugation. Proteins were captured from the supernatant fraction using Ni-NTA agarose beads (Qiagen) (lane E).

3. Results and discussion

3.1. Vector construction

3.1.1. Choice of cloning method. All the cloning methods have given efficiencies of at least 80% in HTP format. Of these, the Gateway recombinatorial cloning method has been the most widely adopted in the SPINE consortium. However, in using the Gateway system it is important to be aware of the effect the *att* recombination sequences may have on expression and/or solubility of the cloned protein if they form part of the translated sequence. In a comparison of Gateway vectors in which several genes were expressed as N-His-tagged constructs either with or without *att* sequences in the translated insert, the Weizmann and Oxford SPINE partners (unpublished) observed that the presence of the *att* sequence was associated with a marked reduction in the level of soluble expression of several proteins (e.g. Fig. 3). This effect can be avoided by positioning the *att* sites outside of the ORF, but some of the flexibility of the system is lost since only a single fusion format is possible. Thus, two vector formats have emerged amongst SPINE groups for using Gateway. Firstly, suites of vectors for generating a variety of fusion proteins have been constructed but with protease cleavage sites to enable removal of the *att* and tag sequence. Secondly, Gateway has been used to insert genes into the vector pDEST14 (Invitrogen) with either short N- or C-terminal His₆ tags via *att* sites immediately downstream of the T7 promoter/enhancer. In this case the RBS is incorporated into the 5' PCR primers used to amplify the target gene as follows.

*att*B1
RBS
Kozak

5' GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAGCCACCATG 3'

This extension would be attached directly to the 5' gene-specific sequence in the case of a C-His-tagged construct. Alternatively, six histidine codons would be added between the extension and the gene-specific sequence to produce an N-His-tagged construct.

Issues with cost, the uncertainty about the effects of the Gateway recombination sequences on expression and the long primers in general required to avoid this have stimulated the development of alternative approaches to HTP cloning. The most widely adopted has been the LIC-PCR system, which has been commercialized by Novagen. The Utrecht group have found that the advantage of this system is that it does not require specialized vectors and reagents are relatively inexpensive, requiring limited amounts of vector (1–3 ng) and insert (1–20 ng) DNA (data from the Utrecht group; Folkers *et al.*, manuscript in preparation). The success rate critically depends on the preparation of high-quality linearized vector which will require batch checking to ensure high efficiency of cloning. A limitation is that one of four bases has to be pre-selected as the 'lock' in the compatible overhangs and hence the base-pair composition of the annealing regions is limited to using the other three bases (Fig. 1). Consequently, the method is not entirely sequence-independent and cannot be used to join any sequence to any other sequence.

The enzyme-free cloning (EFC) method (Tillett & Neilan, 1999; Neilan & Tillett, 2002) has the potential to overcome these limitations. A pairwise comparison of LIC and EFC carried out by the Utrecht group revealed that while for both methods all PCR products ($n = 24$) were successfully cloned, the latter method appeared superior as a larger percentage of the analysed colonies had the correct insert (91% versus 79%; Folkers *et al.*, manuscript in preparation). In-Fusion cloning in addition overcomes the limitation since the enzyme-catalysed reaction is sequence-independent. However, In-Fusion cloning is much less widely used than either Gateway or LIC-PCR at present, although a recent comparison of the two methods concluded that In-Fusion was equally efficient for HTP cloning (Marsischky & LaBaer, 2004). The availability of vectors specifically designed for use with In-Fusion by SPINE partners should increase the utility of the system (Table 2). In comparing the three methods for constructing expression vectors, it is clear that there are advantages and disadvantages to all three. LIC-PCR is the system of choice for minimizing the cost per vector since in contrast to Gateway and In-Fusion it does not require the use of relatively expensive specialized enzyme(s). Only Gateway combines ease of use with maximum flexibility in terms of generating multiple vectors from a single PCR product. In-Fusion cloning probably lies between the other two in terms of unit cost and is the only established HTP to date that is entirely sequence independent.

3.1.2. Choice of fusion tag. HTP protein purification depends on affinity tags to provide a generic strategy. In addition, certain tags have a beneficial effect on protein solubility, especially for the expression of heterologous proteins in *E. coli* (Smith & Johnson, 1988; Kapust & Waugh, 1999). The most widely used format in the SPINE consortium has been the short His₆ tag placed at either the N- or C-terminus of the target gene. In some cases, vectors have been engineered to include a protease cleavage site between the N-terminal tag and the inserted gene to enable removal of the tag post-purification. The most commonly used proteases are from tobacco etch virus (TEV; Parks *et al.*, 1994) and 3C from human rhinovirus (Cordingley *et al.*, 1989). Both enzymes have highly specific linear recognition sequences that are very rarely encountered in other sequences, minimizing the risk of cleavage within the target.

In SPINE, vectors have been constructed using all three cloning strategies described above to produce His-tagged proteins (Table 2). In addition, several groups have developed suites of Gateway-compatible expression vectors that enable different larger fusion proteins to be constructed from a common entry clone (Table 2). The fusion proteins incorporate N-terminal His tags to combine the potential benefits of improved expression levels/solubility of the fusion partner with a generic purification strategy. In specific cases, expression as a fusion protein has rescued otherwise insoluble proteins. For example, the Marseille group found that the hypothetical protein Rv115 from *Mycobacterium tuberculosis* was expressed solubly as a His-maltose-binding protein (MBP) fusion, whereas the His-only version was insoluble. The availability of soluble protein enabled subsequent crys-

tallization and structure solution following removal of the His MBP tag (Canaan *et al.*, 2005).

At present, there is no clear consensus as to which fusion partners give the best performance in terms of enhancing expression/solubility, although the consensus from the literature (including SPINE groups) suggests that MBP and thio-redoxin should be the first choice (Braun *et al.*, 2002; Hammarstrom *et al.*, 2002; Shih *et al.*, 2002; Dyson *et al.*, 2004; Busso *et al.*, 2005). Multiple examples from several SPINE groups, however, have revealed loss of solubility after cleavage, raising doubts on the general usefulness of screening different fusion tags.

3.2. Expression screening

The large-scale and parallel construction of expression vectors either for multiple targets or multiple versions of fewer targets creates a need for parallel expression screening on a small scale. Practical considerations involved in setting up such a screen include the choice of culture conditions (*E. coli* strain, culture volume and media), cell lysis and protein-detection method for soluble expression. For all these variables the key issue is that for a screen to be useful it has to be predictive of the outcome on a larger scale. Amongst SPINE partners different decisions have been made regarding the format of the *E. coli* expression screen. In the accompanying paper (Berrow *et al.*, 2006) the different screening protocols are compared by analysing the results of expression screening a common set of 96 vectors. Here, overall results from applying HTP methods in SPINE are presented.

3.2.1. Solubility screening results. Data from the laboratories of each of the authors were collected *via* a website and analysed (Table 3). It is clear that the technology development described in this report has enabled a very large number of vectors ($n = 3847$) to be constructed for screening expression in *E. coli*. In all laboratories at least two constructs have been made for each gene targeted for structural studies. The majority of these constructs comprised either N- or C-terminal His₆ tags. Each expression experiment has also been carried out on average three times, giving a total of over 10 000 expression trials. The overall result from all the screening experiments was that 32% of the constructs yielded soluble protein, which means that 70% of the targets produced at least one soluble construct (Table 3). Most interestingly, the bacterial and human targets gave similar results, with the viral ones performing slightly worse. These data are comparable to screening results reported by other large-scale structural proteomics projects largely focused on microbial genomes (Christendat *et al.*, 2000; Lesley *et al.*, 2002; Chance *et al.*, 2004).

Follow-on results from the SPINE crystallography groups who used the protocols reviewed in this article indicate that approximately 20% of the soluble proteins identified for scale-up by screening gave diffraction-quality crystals. From these, a total of 150 structures have been solved to date. Details of some of this work are provided in accompanying papers in this volume (*e.g.* Fogg *et al.*, 2006).

Table 3

Summary of expression-screening results.

Data were collected from the laboratories of the authors for activity during the period September 2002 to September 2005 and pooled according to target group (viral, bacterial and human).

Target group	No. of targets	No. of constructs	No. of soluble proteins	Soluble constructs (%)
Viral (<i>e.g.</i> SARS)	234	555	144	26
Bacterial (<i>e.g.</i> <i>Bacillus anthracis</i>)	984	1909	626	33
Human (<i>e.g.</i> kinases, proteases)	497	1383	462	33
Total	1715	3847	1232	32

3.3. Refolding proteins from inclusion bodies

In the study reported by the Marseille group (Vincentelli *et al.*, 2004) 24 proteins that formed inclusion bodies were subjected to a small-scale refolding screen to identify conditions that solubilized the proteins. 17 of 24 remained soluble in at least one of the 96 refolding buffers and 15 were scaled and entered for crystallization trials. Of these, five proteins gave crystal hits (some 20% of the starting set), a notable success rate for this set of targets. It will be of interest to test this screening approach with larger numbers of targets in order to assess the general applicability of the method. In an alternative strategy, the group in Oxford have used a simple generic approach to refolding based on the work of Gao *et al.* (1998), who refolded the extracellular domain of CD8 by rapid dilution into a redox buffer containing arginine. To date, 20 different human and eukaryotic viral proteins have been processed using this protocol. These proteins were selected as either naturally secreted or the ectodomains of cell-surface proteins, all of which formed inclusion bodies following overexpression in *E. coli*. 12 of the proteins (60%) were recovered by refolding as soluble protein and seven have been crystallized (60%), leading to four solved structures (Gao *et al.*, 1998; Mongkolsapaya *et al.*, 1999; Brown *et al.*, 2002; Brown *et al.*, manuscript in preparation; Bahar *et al.*, manuscript in preparation). This indicates that a high success rate (20% of targeted proteins resulting in structures) can be achieved with a single refolding regime for a specific class of protein. However, these examples are limited to secreted proteins and more generally proteins are less amenable to refolding (see, for example, the experiences of Tarbouriech *et al.*, 2006).

3.4. New technology developments

3.4.1. Library methods. Screening is used widely to solve problems where large numbers of variables make it too complex to rationally predict the solution, *e.g.* screening of precipitant conditions in protein crystallization. The design of expression constructs is usually a highly rational process involving careful study of sequence alignments and the use of bioinformatic tools. Within the SPINE project, notable successes in terms of crystallisable protein have resulted from screening multiple constructs from a target protein (see, for example, Berry *et al.*, 2006). Such an approach is operationally

limited, not least by cost, to <100 variants per target. To go much beyond this requires a library-based approach in which thousands of variants can be produced and assayed. Therefore, in this SPINE workpackage both the Grenoble and Stockholm nodes have designed and validated methods in which large numbers of constructs are randomly generated and then tested for production of soluble protein. These methods offer a rescue route for proteins where successful constructs are unable to be generated by standard means. Additionally, once streamlined, library-based screening may become a primary approach, especially for proteins where there is an absence of homologues preventing generation of sequence alignments. The use of automation can greatly improve the reliability by permitting highly repetitive routines such as colony picking whilst aiding sample tracking since these strategies involve the handling of very large numbers of clones (typically 10^3 – 10^5). A representative result from a high-density colony screen is shown in Fig. 4. Once a panel of hits has been isolated from the library, the constructs can be plugged into the standard structural genomics platforms for expression optimization and purification screening using liquid-handling robots.

3.4.2. High-density parallel fermentation. The eight-vessel micro-fermenter system developed by the Pasteur group has been benchmarked with over 50 different *E. coli* clones producing various proteins of *M. tuberculosis* and *M. leprae*. In

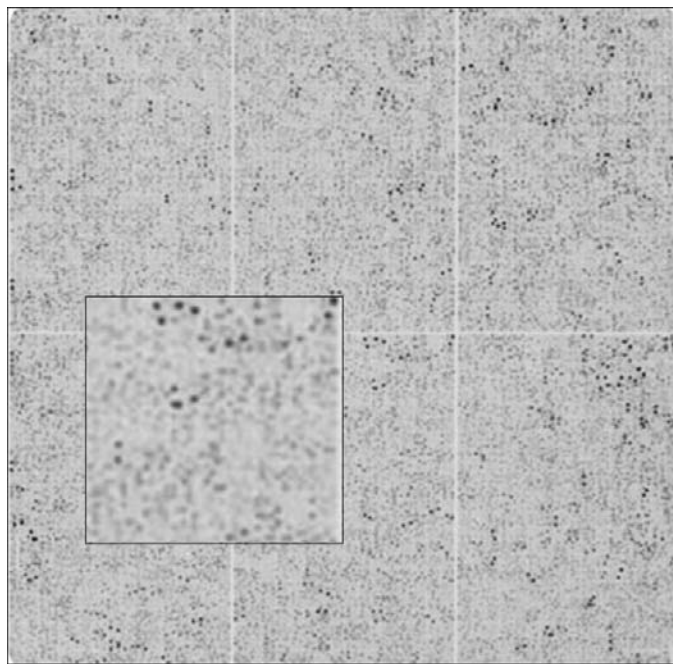


Figure 4

Colony blot of 27 000 clones, arrayed in duplicate, expressing a randomly truncated target gene (code OPTIC 1612). Each clone from the deletion library is printed as an inoculum onto a nitrocellulose filter and grown at 310 K to give the colony array. Expression is induced by transferring the colony array onto LB agar containing IPTG. Following lysis *in situ*, the array is screened for constructs that express soluble proteins. These are detected using a linear peptide fused to the constructs that is efficiently post-translationally modified *in vivo* only if the protein is both soluble and stable (manuscript in preparation).

most cases, the yield of soluble recombinant protein produced in a 70 ml micro-fermenter culture with the high-density medium was as good as or better than the yield obtained in 1 l shake-flask cultures using LB medium. Furthermore, culture protocols developed in micro-fermenters for optimizing recombinant protein production could be reproducibly scaled up when larger quantities were required. Thus, the micro-fermenter battery has been adopted for routine cell culture at the Pasteur Institute and the possibility of making the system more widely available through commercialization is under discussion. A comparable system has been developed by the Novartis Institute and consists of 96×50 ml reaction vessels (Lesley *et al.*, 2002; Lesley & Wilson, 2005). However, the Novartis device does not have the feedback-control systems of the eight-vessel unit, limiting the yield of the system. In addition, the modular design of the Pasteur micro-fermenter would enable a battery of 8×12 units to be assembled to provide a degree of parallel operation equivalent to the Novartis system. Within SPINE, the Stockholm group have also developed parallel fermentation equipment specifically optimized for structural proteomics; this project is detailed in an accompanying paper (Hedrén *et al.*, 2006).

4. Conclusion

Obtaining recombinant proteins in a soluble form suitable for crystallization remains a major bottleneck for HTP structural biology. However, technical advances in cloning and expression are beginning to accelerate protein-production activity. Within the SPINE consortium most laboratories have implemented HTP cloning and expression screening in *E. coli* and this has had a major impact on the ability to process multiple constructs in parallel. As a result, soluble expression of at least domains has been obtained for a relatively high proportion of proteins selected for structural studies. For high-value targets, new methods for high-density screening of libraries of deletion mutants offers a way of identifying sub-regions which express in soluble form in *E. coli* without prior knowledge of domain organization. Overall, the HTP production of recombinant proteins in *E. coli* remains central to increasing the rate of protein structure solution in a cost-effective manner.

We thank all our colleagues for their contributions to the work reviewed in this article. This project is funded by the European Commission as SPINE (Structural Proteomics In Europe) Contract No. QLG2-CT-2002-00988 under the Integrated Programme ‘Quality of Life and Management of Living Resources’.

References

- Albeck, S., Burstein, Y., Dym, O., Jacobovitch, Y., Levi, N., Megeed, R., Michael, Y., Peleg, Y., Prilusky, J., Schreiber, G., Silman, I., Unger, T. & Sussman, J. L. (2005). *Acta Cryst.* **D61**, 1364–1372.
- Aslandis, C. & deJong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.

- Berrow, N. S. *et al.* (2006). *Acta Cryst.* **D62**, 1218–1226.
- Berry, I. M., Dym, O., Esnouf, R. M., Harlos, K., Meged, R., Perrakis, A., Sussman, J. L., Walter, T. S., Wilson, J. & Messerschmidt, A. (2006). *Acta Cryst.* **D62**, 1137–1149.
- Betton, J. M. (2004). *Biochimie*, **86**, 601–605.
- Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M. & LaBaer, J. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 2654–2659.
- Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
- Busso, D., Delagoutte-Busso, B. & Moras, D. (2005). *Anal. Biochem.* **343**, 313–321.
- Busso, D., Poussin-Courmontagne, P., Rose, D., Ripp, R., Litt, A., Thierry, J.-C. & Moras, D. (2005). *J. Struct. Funct. Genomics*, **6**, 81–88.
- Canaan, S., Sulzenbacher, G., Roig-Zamboni, V., Scappuccini-Calvo, L., Frassinetti, F., Maurin, D., Cambillau, C. & Bourne, Y. (2005). *FEBS Lett.* **579**, 215–221.
- Chance, M. R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J. E., Radhakannan, T. & Marinkovic, N. (2004). *Genome Res.* **14**, 2145–2154.
- Christendat, D. *et al.* (2000). *Nature Struct. Biol.* **7**, 903–909.
- Cordingley, M. G., Register, R. B., Callahan, P. L., Garsky, V. M. & Colonno, R. J. (1989). *J. Virol.* **63**, 5037–5045.
- Cornvik, T., Dahlroth, S. L., Magnosdottir, A., Herman, M. D., Knaust, R., Ekberg, M. & Nordlund, P. (2005). *Nature Methods*, **2**, 507–509.
- Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. (2004). *BMC Biotechnol.* **4**, 32.
- Fogg, M. J. *et al.* (2006). *Acta Cryst.* **D62**, 1196–1207.
- Folkers, G. E., van Buuren, B. N. & Kaptein, R. (2004). *J. Struct. Funct. Genomics*, **5**, 119–131.
- Gao, G. F., Gerth, U. C., Wyer, J. R., Willcox, B. E., O'Callaghan, C. A., Zhangm Z., Jones, E. Y., Bell, J. I. & Jakobsen, B. K. (1998). *Protein Sci.* **7**, 1245–1249.
- Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. & Hard, T. (2002). *Protein Sci.* **11**, 313–321.
- Hart, D. & Tarendeau, F. (2006). *Acta Cryst.* **D62**, 19–26.
- Hartley, J. L., Temple, G. F. & Brasch, M. A. (2000). *Genome Res.* **10**, 1788–1795.
- Hartman, S., Trinklein, N. D., Anton, E. D., Marticke, S. S., Nguyen, L. & Myers, R. M. (2005). *Clonetechniques*, Jan. 2005. http://www.clontech.com/clontech/archive/JAN05UPD/tech_notes_cloning.shtml.
- Haun, R. S., Serventi, I. M. & Moss, J. (1992). *Biotechniques*, **13**, 515–518.
- Hedrén, M., Ballagi, A., Mörtzel, L., Rajkai, G., Stenmark, P., Sturesson, C. & Nordlund, P. (2006). *Acta Cryst.* **D62**, 1227–1231.
- Kapust, R. B. & Waugh, D. S. (1999). *Protein Sci.* **8**, 1668–1674.
- Knaust, R. K. & Nordlund, P. (2001). *Anal. Biochem.* **297**, 79–85.
- Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
- Lesley, S. A. & Wilson, I. A. (2005). *J. Struct. Funct. Genomics*, **6**, 71–79.
- Marsischky, G. & LaBaer, J. (2004). *Genome Res.* **14**, 2020–2028.
- Mongkolsapaya, J., Grimes, J. M., Chen, N., Xu, X. N., Stuart, D. I., Jones, E. Y. & Sreaton, G. R. (1999). *Nature Struct. Biol.* **6**, 1048–1053.
- Neilan, B. A. & Tillett, D. (2002). *Methods Mol. Biol.* **192**, 125–132.
- Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.
- Quevillon-Cheruel, S. *et al.* (2003). *J. Synchrotron Rad.* **10**, 4–8.
- Quevillon-Cheruel, S., Liger, D., Leulliot, N., Graille, M., Poupon, A., de La Sierra-Gallay, I. L., Zhou, C. Z., Collinet, B., Janin, J. & van Tilbeurgh, H. (2004). *Biochimie*, **86**, 617–623.
- Shih, Y. P., Kung, W. M., Chen, J. C., Yeh, C. H., Wang, A. H. & Wang, T. F. (2002). *Protein Sci.* **11**, 1714–1719.
- Smith, D. B. & Johnson, K. S. (1988). *Gene*, **67**, 31–40.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Tarbouriech, N., Buisson, M., Géoui, T., Daenke, S., Cusack, S. & Burmeister, W. P. (2006). *Acta Cryst.* **D62**, 1276–1285.
- Tepliyakov, A., Obmolova, G., Bir, N., Reddy, P., Howard, A. J. & Gilliland, G. L. (2003). *J. Struct. Funct. Genomics*, **4**, 1–10.
- Thao, S., Zhao, Q., Kimball, T., Steven, E., Blommel, P. G., Ritters, M., Newman, C. S., Fox, B. G. & Wrobel, R. L. (2004). *J. Struct. Funct. Genomics*, **5**, 267–276.
- Tillett, D. & Neilan, B. A. (1999). *Nucleic Acids Res.* **27**, e26.
- 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. (2004). *J. Struct. Funct. Genomics*, **5**, 195–204.
- Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V. & Cambillau, C. (2003). *Acc. Chem. Res.* **36**, 165–172.
- Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappuccini-Calvo, L., Bourne, Y., Cambillau, C. & Bignon, C. (2004). *Protein Sci.* **13**, 2782–2792.
- Vincentelli, R., Canaan, S., Offant, J., Cambillau, C. & Bignon, C. (2005). *Anal. Biochem.* **346**, 77–84.
- Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. & Vidal, M. (2000). *Methods Enzymol.* **328**, 575–592.