

Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study

Nick S. Berrow,^a K. Büssov,^b
B. Coutard,^c J. Diprose,^a
M. Ekberg,^d G. E. Folkers,^e
N. Levy,^f V. Lieu,^d R. J. Owens,^a
Y. Peleg,^f C. Pinaglia,^g
S. Quevillon-Cheruel,^g L. Salim,^h
C. Scheich,^b R. Vincentelli^c and
Didier Busso^{h*}

^aOxford Protein Production Facility, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, England, ^bProtein Structure Factory, Heubnerweg 6, 14059 Berlin, Germany, ^cArchitecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS/Universités de Provence/Université de la Méditerranée, Parc Scientifique et Technologique de Luminy, 163 Avenue de Luminy Case 932, 13288 Marseille CEDEX 09, France, ^dDepartment of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden, ^eBijvoet Center for Biomolecular Research, NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, ^fThe Israel Structural Proteomics Center, Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel, ^gInstitut de Biochimie et de Biophysique Moléculaire et Cellulaire, UMR 8619, Bâtiment 430, Université de Paris-Sud, 91405 Orsay CEDEX, France, and ^hInstitut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, BP 163, 67404 Illkirch CEDEX, France

Correspondence e-mail:
djbusso@titus.u-strasbg.fr

Producing soluble proteins in *Escherichia coli* is still a major bottleneck for structural proteomics. Therefore, screening for soluble expression on a small scale is an attractive way of identifying constructs that are likely to be amenable to structural analysis. A variety of expression-screening methods have been developed within the Structural Proteomics In Europe (SPINE) consortium and to assist the further refinement of such approaches, eight laboratories participating in the network have benchmarked their protocols. For this study, the solubility profiles of a common set of 96 His₆-tagged proteins were assessed by expression screening in *E. coli*. The level of soluble expression for each target was scored according to estimated protein yield. By reference to a subset of the proteins, it is demonstrated that the small-scale result can provide a useful indicator of the amount of soluble protein likely to be produced on a large scale (*i.e.* sufficient for structural studies). In general, there was agreement between the different groups as to which targets were not soluble and which were the most soluble. However, for a large number of the targets there were wide discrepancies in the results reported from the different screening methods, which is correlated with variations in the procedures and the range of parameters explored. Given finite resources, it appears that the question of how to most effectively explore 'expression space' is similar to several other multi-parameter problems faced by crystallographers, such as crystallization.

1. Introduction

Small-scale screening for soluble expression in *Escherichia coli* is a key feature of the experimental pipelines that have been implemented for structural proteomics projects in a number of European laboratories (Alzari *et al.*, 2006). The justification for such screening is that it provides information to guide subsequent decisions on whether to invest in large-scale purification of a given construct. Reducing the time, effort and cost of each expression trial should enable more constructs to be tested and expression parameters to be optimized (Folkers *et al.*, 2004 and references therein). However, screening methods are only of value if they satisfy two fundamental criteria. Firstly, they must be reproducible and secondly, they must give reliable qualitative and ideally quantitative predictions of the outcome of the larger scale protein production needed to produce sufficient protein for structural studies. The central question is therefore 'how should we configure small-scale expression screening in order to establish a route to the production of milligram amounts of soluble protein on scale-up?' To this end, we have compared the methods routinely used in eight different laboratories in the SPINE consortium (Berlin, Marseille, Orsay, Oxford,

Received 23 March 2006

Accepted 9 August 2006

Table 1
Variable parameters at key stages of the expression screen process by centre.

Parameter	Berlin	Marseille	Orsay	Oxford	Stockholm	Strasbourg	Utrecht	Weizmann
<i>E. coli</i> strain(s)	Rosetta (DE3)	BL21 (DE3)-pLysS, Rosetta (DE3)-pLysS, Origami (DE3)-pLysS, C41 (DE3)-pLysSRARE	BL21 (DE3)-Gold, Rosetta (DE3)	B834 (DE3), Rosetta (DE3)-pLysS	BL21 (DE3), RosettaII (DE3)	BL21 (DE3)	BL21 (DE3), RosettaII (DE3)	BL21 (DE3), BL21-AI
96 transformations in parallel	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Plating format	2 × 48-well plate	4 × 24-well plate (as reserve)	No plating	4 × 24-well plate	4 × 24-groove plate	16 × 6-well plates	8 × 12-well plates	Individual plates
Pre-culture source	Plated colony	Transformation mix	Transformation mix	Plated colony	Plated colony	Plated colony	Plated colony	Plated colony
Pre-culture format/volume (µl)	96-well deep-well plate/100	96-well deep-well plate/500	24-well deep-well plate/3000	96-well deep-well plate/500	96-well deep-well plate/1000	96-well deep-well plate/500	96-well plate/100	14 ml tubes/3000
Pre-culture media	2YT+2% (w/v) glucose	LB	2YT	GS96+1% (w/v) glucose	LB	LB	LB	LB
Pre-culture temperature (K)	310	310	310	310	310	310	310	310
Pre-culture duration (h)	18	18	18	18	18	18	3–4	18
Expression-culture format/volume (µl)	96-well deep-well plate/1000	96-well deep-well plate/1000	24-well deep-well plate/3000	24-well deep-well plate/2500	96-well deep-well plate/1000	24-well deep-well plate/2000	96-well plate/100	14 ml tubes/3000
Expression-culture media	TB/SB+KPB	SB/TB/2YT	2YT	GS96+1% (w/v) glucose or Overnight Express Instant TB	LB	ZYM-5052	LB	LB
Expression-culture temperature (phase 1 growth/pre-induction) (K)	310	310	310	310	310	310	310	310
Control point/induction point	2 h	2 h	OD _{600 nm} = 0.8–1.2	OD _{600 nm} = 0.6 (IPTG-induced), 3 h (auto-induced)	2.5 h at 298 K or 5.5 h at 288 K	OD _{600 nm} = 1–2	OD _{600 nm} = 0.8, 1.6, 2.3 and 3	OD _{600 nm} = 0.6–0.8
Induction method	IPTG (1.0 mM)	IPTG (1.0 mM)	IPTG (0.5 mM)	IPTG (0.5 mM) or auto-induction	IPTG (0.2 mM)	Auto-induction	IPTG (1.0 mM)	IPTG (0.05 mM) or IPTG (0.05 mM) + 0.02% (w/v) arabinose
Expression culture temperature (phase 2 expression/post-induction) (K)	293, 301 and 310	290, 298 and 310	288 and 310	293 (IPTG-induced), 298 (auto-induced)	288 and 298	293	293 and 310	303
Harvest point	3 h (301 and 310 K), 6 h (293 K)	3 h (310 K), 18 h (290 K and 298 K)	3 h (310 K), 18 h (288 K)	18 h (IPTG-induced), 24 h (auto-induced)	3 h (298 K), 18 h (288 K)	18 h	OD _{600 nm} = 3.0 (~12 h)	3 h (303 K)
Harvest method	Centrifugation	Centrifugation	Centrifugation	Centrifugation	Centrifugation	Centrifugation	N/A (see below)	Centrifugation
Lysis buffer	20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 mM EDTA, 10 mM imidazole, 1 mM PMSF, 0.1% Brij58, 380 mg ml ⁻¹ lysozyme	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM PMSF, 0.2 mg ml ⁻¹ lysozyme	20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol, benzonase	50 mM NaH ₂ PO ₄ pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% (v/v) Tween-20	50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 U ml ⁻¹ benzonase, 1.0 mg ml ⁻¹ lysozyme, 0.015% (w/v) dodecylmalto-side, protease inhibitors	50 mM Tris-HCl pH 7.5, 250 mM NaCl, 10% glycerol, 2.5 mM β-mercaptoethanol	PopCulture, lysozyme, benzonase added directly to culture	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM PMSF, protease inhibitors

Table 1 (continued)

Parameter	Berlin	Marseille	Orsay	Oxford	Stockholm	Strasbourg	Utrecht	Weizmann
Cell-disruption method	Freeze/thaw	Freeze/thaw and sonication	Freeze/thaw and sonication	Freeze/thaw	Freeze/thaw	Sonication	N/A (see above)	Sonication
Fractionation method	Total sample processed by metal-chelate protocol	Total sample processed by metal-chelate protocol	Filtration or centrifugation (1 h at ~5000g)	Centrifugation (30 min at ~5000g)	Centrifugation (30 min at ~2000g)	Centrifugation (1 h at ~3500g)	N/A (see above)	Centrifugation (15 min at ~15000g)
Screening/partial purification method	Metal chelate (Ni-NTA SuperFlow/positive-pressure filtration)	Metal chelate (nickel-Sepharose Fast Flow/vacuum filtration)	Metal chelate (Ni-NTA)	Metal chelate (Ni-NTA magnetic beads)	Metal chelate (Ni-NTA magnetic beads)	Metal chelate (cobalt-TALON/vacuum filtration)	Metal chelate (Ni-NTA magnetic beads)	Metal chelate (Ni-NTA agarose beads)
Visualization/quantification method	SDS-PAGE/manual comparison to standards	Automated immuno/dot-blotting <i>versus</i> standard curve supplemented with SDS-PAGE/manual comparison to standards	SDS-PAGE/manual comparison to standards	SDS-PAGE/manual comparison to standards	Automated immuno/dot-blotting <i>versus</i> standard curve supplemented with SDS-PAGE/manual comparison to standards	SDS-PAGE/manual comparison to standards	SDS-PAGE/manual comparison to standards	SDS-PAGE/manual comparison to standards
Amount of material loaded for analysis (in fraction of the culture volume)	1/112th for total proteins and 1/12th for eluted proteins on SDS-PAGE	1/3rd for the immuno/dot-blotting and 1/16th for SDS-PAGE for eluted proteins	1/100th for total and soluble proteins on SDS-PAGE	1/20th for the SDS-PAGE of eluted proteins	1/2.5th for eluted proteins on SDS-PAGE and 1/12.5th of total soluble and eluted proteins for immuno/dot-blotting	1/10th of total and soluble proteins on SDS-PAGE	1/6th of total and soluble proteins on SDS-PAGE	1/5th of insoluble and eluted and 1/62nd of soluble proteins on SDS-PAGE

Stockholm, Strasbourg, Utrecht and the Weizmann) for soluble expression screening in *E. coli* using a common set of 96 expression vectors. These encode proteins ranging in molecular weight from 9 kDa to more than 100 kDa (see supplementary material¹) and were largely eukaryotic or viral in origin. The parameters varied included *E. coli* strain, growth temperature, optical density at induction, culture-vessel size and design, agitation levels, media and lysis method. We present simple statistical analyses of the results obtained in a single representative screening experiment performed by each centre and suggest guidelines for the future refinement of such protocols.

2. Materials and methods

General methods are described below and specific features or variants of the methodologies for the different laboratories are specified in Table 1.

2.1. Target vectors

The test set of 96 expression constructs was assembled from seven of the eight SPINE groups that participated in the study, with each group contributing between eight and 16 plasmids (as mini-preparations). The large majority of these targeted human or viral proteins. The choice of targets was not constrained and only the vector details, molecular weight and

other biophysical properties expected of the expressed products (including fusion partners) were submitted (see supplementary material). It is clear from the results that the panel of targets contained a significant number of plasmids that had failed to yield soluble protein in-house and the set is therefore representative of 'difficult' targets. All plasmids used the T7 promoter system for transcriptional regulation in combination with *E. coli* strains harbouring the DE3 prophage (Studier *et al.*, 1990). In addition, all constructs encoded a His₆ tag fused to the protein of interest to enable routine purification using immobilized metal-chelating resin (see supplementary material). The constructs were transformed (at a single site, Oxford) in parallel into Omnimax bacteriophage-resistant competent cells in 96-well format (Invitrogen, Paisley, Scotland) for plasmid propagation. After culturing in 96-well blocks, the plasmids were prepared on a Biorobot 8000 using the 96 Turbo Miniprep kit (Qiagen, West Sussex, England). These plasmid preparations, of a standard concentration and quality, were then distributed in 96-well format to the participating SPINE centres for use in expression screen trials.

2.2. Expression protocol

2.2.1. Host strains. One or more *E. coli* strains drawn from a panel comprising the *lon* and *ompT* protease-deficient BL21 (DE3) and derivatives [including Rosetta (DE3) and RosettaII (DE3), B834 (DE3), BL21-Gold (DE3) and BL21-AI] were used by each of the laboratories. The Rosetta (Merck) strains carry a chloramphenicol-resistant plasmid,

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5098). Services for accessing this material are detailed at the back of the journal.

pRARE, that contributes tRNAs for codons rarely used in *E. coli*. LysS and LysE variants of the pRARE plasmid constitutively express T7 lysozyme, a natural inhibitor of T7 RNA polymerase activity, reducing polymerase activity in uninduced cells. The B834 (DE3) strain (Merck) is the methionine-auxotrophic version of BL21 (DE3), which allows efficient selenomethionine labelling of expressed proteins (Hendrickson *et al.*, 1990). The BL21-Gold (DE3) strain (Stratagene, La Jolla, CA, USA) has been developed to increase transformation efficiency and is *recA*-deficient to improve the stability of expression vectors. BL21-AI (Invitrogen, Carlsbad, CA, USA) carries a chromosomal copy of the T7 RNA polymerase gene under the control of the arabinose-inducible *araBAD* promoter conferring low basal expression prior to induction and dose-dependent induction. Finally, C41 (DE3) (Avidis, St Beauzire, France) is a mutant that allows overexpression of proteins that are toxic to the BL21 (DE3) parental strain (Dumon-Seignovert *et al.*, 2004). In this study, the LysS variant of the pRARE plasmid was transformed into the C41 (DE3) strain by Marseille. The only non-BL21 (DE3) derived strain used was the K-12 based Origami strain (Merck) which carries mutations in both the thioredoxin reductase (*trxB*) and the glutathione reductase (*gor*) genes, which enhance disulfide-bond formation in the cytoplasm of *E. coli*.

Competent cells of the expression strains were prepared by each group using variations on the calcium chloride procedure (Hanahan, 1983; Inoue *et al.*, 1990; Nakata *et al.*, 1997), except for Stockholm where Z-competent cells (Zymo Research Corporation, Orange, CA, USA) were used. Transformations were performed in parallel in either 96-well plates or racked tubes in 96-well format, except at the Weizmann where individual tubes and plates were used. No major differences in transformation efficiencies were noted between the centres in spite of the variations between the protocols used (volume of competent cells used, recovery volumes and plating volumes). Typically, the transformation mix was plated out and a single colony was picked for the starter culture prior to dilution and regrowth for the expression-screening experiment. To test reproducibility, the Berlin centre inoculated two pre-cultures, each from one isolated colony, and subsequently obtained similar expression results (data not shown). Three centres (Marseille, Orsay and Utrecht) directly inoculated the pre-culture from the transformation mix (to save time and circumvent the error-prone colony-picking step) without impairing growth or expression (data not shown).

2.2.2. Culture conditions. To reduce variations in culture density between clones during the growth step, cultures were inoculated from starter cultures which had reached saturation following growth overnight at 310 K. One centre (Utrecht) did not use a pre-culture as cultures were monitored throughout the entire growth phase and all cultures were then induced at identical OD. The dilution factor for inoculation (pre-culture to expression culture) varied between laboratories. Either a volume was added to give a specific OD at inoculation (usually 0.1 OD_{600nm}) or a fixed dilution of the starter culture was used (from 1/10th to 1/250th of the culture volume).

With one exception (Weizmann, where shaken 14 ml tubes were used), cultures were performed in parallel using shaken deep-well blocks: either four 24-deep-well blocks or a single 96-deep-well block. A variety of media were used (LB, 2YT, TB/SB+KPB, GS96 and non-inducing ZYM-5052 and Overnight Express Instant TB); media formulations are given in the supplementary material). Cultures were grown at various temperatures and shaking speeds either for a fixed time or until a given OD_{600nm} was obtained before induction (see Table 1).

2.3. Protein expression/solubility analysis

The cells were harvested by centrifugation and resuspended in an appropriate volume of lysis buffer, except in Utrecht where PopCulture solution (Merck) was used to lyse the cells whilst still suspended in medium. Lysis buffer composition differed from one laboratory to another, but generally consisted of a buffered saline solution (100–500 mM NaCl in 20–200 mM Tris-HCl or phosphates at pH 7.5 or 8) containing additives such as β -mercaptoethanol, PMSF, glycerol, DNase I, benzonase, lysozyme and various detergents (Table 1). Cells were disrupted by either mechanical means (sonication by a probe modified for plate-based sonication) or by the action of lysozyme with or without freeze-thaw cycle(s). Five laboratories (Orsay, Oxford, Stockholm, Strasbourg and Weizmann) took post-lysis samples to measure the total cell protein content (see supplementary material). The lysate was separated, either by centrifugation or filtration, into soluble fractions (supernatant or filtrate) or insoluble fractions (pellet or retentate). The remaining centres (Berlin, Marseille, Utrecht) proceeded to the affinity mini-purification step without clarification of the lysate.

Soluble proteins were either analysed directly by SDS-PAGE or purified by immobilized metal-affinity chromatography (IMAC) prior to SDS-PAGE analysis. Matrix-bound proteins were eluted by the addition of either SDS-PAGE sample/Laemmli buffer or an elution buffer containing high concentrations of imidazole (>200 mM). Two centres (Marseille and Stockholm) used dot-blotting to screen for expression/solubility and confirmed profiles by SDS-PAGE (Vincentelli *et al.*, 2005; Knaust & Nordlund, 2001; Cornvik *et al.*, 2005). Briefly, the technique consists of applying the different fractions (total or insoluble protein, soluble protein and histidine-chelate-purified protein) directly onto PVDF/nitrocellulose membrane by filtration under vacuum. Following immobilization, the membranes were thoroughly washed and the His₆-tagged proteins detected by tag-specific antibodies.

2.4. Statistical analysis

Each group (*j*) scored the level of soluble expression obtained for each target *k* (SOL_{*j,k*}) using their standard detection methods (SDS-PAGE and/or dot-blot) in their routine screen according to a standardized regime (−1, not performed; 0, no detectable expression or degraded protein; 1,

expression predicted to give less than 0.5 mg l⁻¹ at scale; 2, 0.5–5.0 mg l⁻¹; 3, >5 mg l⁻¹).

Several laboratories investigated expression under multiple growth conditions for each target using their standard procedure for small-scale expression screening and for some sites these results were consolidated and the best-case score for each target submitted, whereas other laboratories submitted multiple sets of scores, each derived from an individual growth condition, which were similarly consolidated for certain analyses. The best-case score for target *k* in laboratory *j*, BEST_{*j,k*}, is simply the greatest value of SOL_{*j,k*}. In all cases, the scores supplied by each group came from a single screening experiment and were not further moderated.

Scores for each target, *k*, were summed across laboratories (*j*) to give an aggregate score TOT_TARGET_{*k*},

$$\text{TOT_TARGET}_k = \sum_j \text{BEST}_{j,k},$$

and, for each laboratory, across targets to give TOT_LAB_{*j*},

$$\text{TOT_LAB}_j = \sum_k \text{BEST}_{j,k}.$$

The scores were further binned in two separate ways, firstly as to whether there was a consensus that soluble expression was achievable even at a low level and secondly to identify those targets where there was a consensus that soluble expression was at a sufficient level to be useful for scale-up. The first measure, EXPRESSION_{*k*}, was defined as positive if the majority of groups reported scores of 1–3 for BEST_{*j,k*}. The second measure, SCALE_{*k*}, was defined as positive if the majority of groups scored BEST_{*j,k*} at 2 or 3.

3. Results and discussion

3.1. Reproducibility and predictive power

To justify the value of the detailed analysis of the comparative data which form the heart of this study, we firstly present additional data to demonstrate that such protocols can be both reproducible and provide reliable predictions of soluble protein yield on scale-up.

3.1.1. Reproducibility. Evidence for the reproducibility of small-scale screening comes from previous (unpublished) results from Oxford where, in a separate study, 33 out of 66 constructs produced soluble protein, with 85% consistency between two separate experiments. The Berlin group evaluated expression from two independent clones (see §2.2.2) with consistent results (data not shown). To extend this to address the question of reproducibility between different laboratories, the results of soluble expression for the first 24 targets of the benchmark list (see supplementary material) will be considered. The results of screening these targets by the Utrecht, Oxford, Stockholm and Marseille groups are shown in Fig. 1 and summarized in Table 2. The Utrecht group observed that 14 of the first 24 targets were expressed (Fig. 1*a*) and that nine of these showed soluble expression (Fig. 1*b*, Table 2). The Oxford screen detected eight of these nine proteins as soluble and gave a further three soluble hits

Table 2

The soluble expression scores assigned by the Utrecht, Oxford, Stockholm and Marseille groups for the each of the targets 1–24 based on the results shown in Fig. 1.

The expected molecular weight of each of the targets is also given. A full set of data for all targets and groups are given in the supplementary material.

Target	MW (kDa)	Utrecht	Oxford	Stockholm	Marseille
1	21.75	2	1	0	3
2	15.6	0	1	0	0
3	19.45	3	3	3	3
4	12.3	3	3	3	3
5	25	0	0	0	0
6	24.9	0	0	0	0
7	10.55	2	1	3	3
8	22.4	0	0	0	3
9	11.8	0	0	0	0
10	13.9	0	0	0	0
11	13.7	3	1	0	1
12	13.6	2	0	0	1
13	22.6	2	2	3	1
14	13.15	3	3	3	3
15	34.75	2	1	2	0
16	19.9	0	0	0	0
17	12.00	0	1	0	3
18	13.60	3	2	3	3
19	30.00	0	0	0	0
20	15.70	0	0	0	0
21	9.80	0	0	0	0
22	21.30	0	0	0	0
23	11.70	0	0	0	0
24	51.90	0	0	0	1

(targets 2, 15 and 17, Fig. 1*c* and Table 2). This shows that there is consistency between the two groups using a similar protocol, but also a difference in the overall hit rate. The differences between the two groups are most likely to be the result of the smaller culture volume used by the Utrecht group, which prevented detection of the weakly expressed clones. The results of using dot-blot to detect soluble expression for the first 24 targets are shown in Figs. 1(*d*) and 1(*e*) and summarized in Table 2. The Stockholm group scored seven proteins as soluble, whilst soluble expression of 12 proteins was obtained in the Marseille screen (Fig. 1*e* and Table 2). However, the Marseille screen explored a much larger range of expression conditions compared with the Stockholm protocol (Table 1). A comparison of the results obtained by SDS-PAGE with the dot-blot screens shows that six of the constructs identified as soluble by SDS-PAGE also gave visible signals in both dot-blot screens (Table 2). For the most part, these were targets with the highest soluble expression scores. Conversely, all four groups were generally in agreement upon the targets which did not show any soluble expression. Thus, it appears that the detection of targets that are either highly soluble or insoluble is reproducible between different groups. Variations appear for those proteins that are only expressed with relatively limited solubility. This observation is borne out by the analysis of data on soluble expression from all the groups (see below).

3.1.2. Predictive value of the small-scale expression screen.

For 31 of the 96 constructs in the benchmark study, the yields of soluble protein obtained following scale-up of cultures to at least 1 l culture volume were available. Larger scale cultures

were carried out using the culture condition identified at small scale that gave the best soluble expression score. By comparing these results with the level of soluble expression estimated from the small-scale screen, some assessment of the predictive value of screening can be made. Of the 31 targets, only three appeared as 'non-predictive' outliers. Construct 1 was scored as 1 but yielded $>5 \text{ mg l}^{-1}$ at scale, whereas constructs 17 and 18 scored 3 but failed to produce soluble protein on scale-up (although in these cases the scaled expression cultures were performed in minimal media, probably accounting for the discrepancy). In general, the comparison supports the predictive value of screening for soluble expression on a small scale, which enables candidates for scale-up to be ranked. For routine scale-up of *E. coli* cultures, 1 l cultures are typical, particularly for producing multiple targets in parallel for structural genomics projects. With the advent of sub-microlitre volumes of drops for crystallization screens (Sulzenbacher *et al.*, 2002; Walter *et al.*, 2005), a 1 l culture can be expected to yield sufficient material for a primary crystallization screen (which ranges from one set of 96 conditions up to five 96-condition plates, depending upon the centre), since with 100 nl protein drops only 150 μg of concentrated protein (typically at 10 mg ml^{-1}) is required for each 96-well plate of screening conditions.

The present analysis of 31 targets (which are probably representative of relatively difficult targets) suggests that triage on the basis of the small-scale results would dramatically reduce the number of targets that progress to large-scale culture preparations. The disadvantage of the screening approach is the potential to 'miss' a small percentage of proteins by not progressing to scale-up cultures. Nevertheless, on the basis of these results we find the loss of a small number of targets for scale-up is more than balanced by the increased throughput possible with small-scale expression screening experiments, which allows a greater range of constructs to be explored.

3.2. Overall success levels

The results of expression analysis are detailed in the supplementary material and summarized in Fig. 2. Overall, detectable soluble expression ($\text{BEST} > 0$) was reported for 81 of the 96 targets by at least one laboratory. However, only 25 of these achieved consensus expression across the laboratories (EXPRESS_j positive), eight of which were predicted as scaleable (SCALE_j positive). There was a dramatic variation in the number of proteins reported as expressed by the different laboratories, ranging from 56 reported by Berlin to 17 by Strasbourg. It is perhaps not surprising that the number from Strasbourg should be low since they were the only laboratory to use a single procedure (Busso *et al.*, 2005), whereas other laboratories tested several different variables, as discussed below. Not only was there a variation in the overall number of targets reported as expressing, but there was considerable variation as to which targets were successful in each laboratory. This was reflected in the fact that pairwise linear correlation coefficients calculated on the scores (BEST)

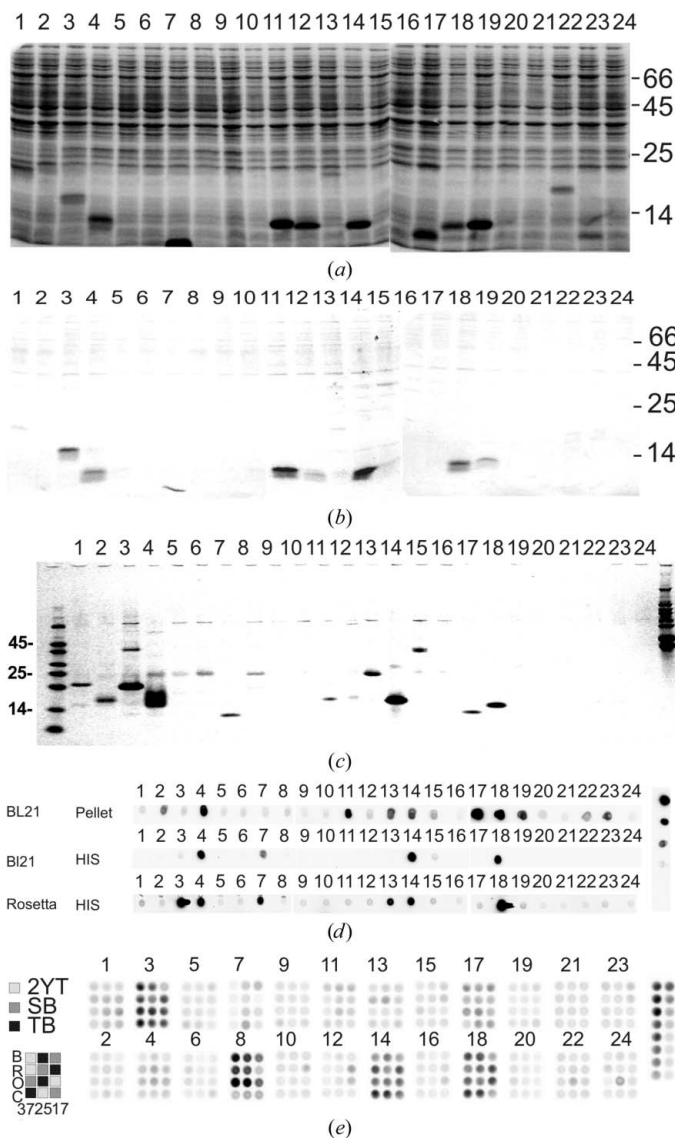


Figure 1

Representative experiments of total expression (a) and soluble expression after IMAC purification (b) from Utrecht using RosettaII (DE3), as determined by SDS-PAGE. (c) Soluble expression from Oxford after IMAC in Rosetta (DE3)-pLysS as determined by SDS-PAGE. The first and last lanes are molecular-weight markers. (d) Insoluble (pellet) and soluble expression from Stockholm after IMAC in the indicated bacterial strains BL21 (DE3) or RosettaII (DE3). Protein was detected using the dot-blot procedure described by Knaust & Nordlund (2001). The scale on the right is displayed as an indicator for 'very good', 'good', 'weak', 'poor' and 'non' expressed protein. (e) Soluble expression as determined by Marseille using a dot-blot procedure (Vincentelli *et al.*, 2005), where expression was performed in parallel (1152 conditions) using three different media (light grey, 2YT; dark grey, SB; black, TB), three different culture temperatures (290, 298 and 310 K) and four different bacterial strains [B = BL21 (DE3)-pLysS, R = Rosetta (DE3)-pLysS, O = Origami (DE3)-pLysS, C = C41 (DE3)-pLysSRARE]. 12 out of the 36 possible conditions were used in an incomplete factorial approach as schematically indicated in the shaded table [e.g. the first spot of every expression test refers to an experiment performed at 310 K in BL21 (DE3)-pLysS using 2YT]. The results for standard amounts spotted for referencing are shown alongside (left column from top to bottom, 2000, 1500, 1000, 900, 800, 700, 600 and 500 ng per dot; right column from top to bottom, 400, 300, 200, 100, 50, 25, 12.5 and 0 ng per dot). The quantification is performed automatically with the microplate reader implemented on the Tecan robot (photon, calibration curve). Molecular-weight markers in kDa are indicated.

between laboratories did not reveal any convincing relationships. Analysis of variance indicated that both site (k) and target (j) contributed significant variance to $BEST_{jk}$ ($p < 0.05$), but it was not possible to split the results into coherent groupings. Where a consensus was seen (42% of the targets), the target generally either expressed well ($BEST = 3$) or not at all ($BEST = 0$). However, for a substantial number of the targets the result reported was dependent on the parameters explored by the small-scale screen. These parameters are considered in more detail below.

3.3. Comparison of screen parameters

The number of parameters tested in the primary screen varied considerably between groups. For example, Strasbourg reported the results from a single condition, whereas Marseille sampled three different variables (strain, medium and temperature) in an incomplete factorial approach where 12 conditions derived from a sparse matrix covering 36 combinations are used (adapted from Abergel *et al.*, 2003). Most groups varied temperature and used at least two expression strains. The detection methods were more standard with groups using SDS-PAGE and/or dot-blot analyses.

3.3.1. Aeration/media. Berlin reported expression ($BEST > 0$) for the largest number of targets. The distinctive feature of their protocol was very high-speed agitation of the 1 ml culture in 96-deep-well growth plates to ensure good aeration (1200 rev min⁻¹ with a 2 mm orbit), carried out with an enriched and buffered media (TB/SB+KPB, see supple-

mentary material and Scheich *et al.*, 2003), which was also used by Marseille and Oxford. The Marseille group (Vincentelli *et al.*, 2003) observed that whatever the medium used, the number of hits for a given strain at a given temperature were comparable, suggesting that the media were not a major determinant of protein expression/solubility. The enhanced expression achieved by Berlin using high aeration is noteworthy, but may not be predictive of expression levels attainable in standard shaker-flask-based scale-up protocols.

3.3.2. Culture temperature. The Berlin results also point to the culture temperature following IPTG induction affecting the production of soluble protein. The overall sum of soluble expression scores (TOT_LAB_{Berlin}) was 64 for proteins expressed at 310 K compared with 119 at 293 K ($BEST > 0$ for 27 and 50 targets at 310 and 293 K, respectively). The Marseille group observed similar behaviour where (for a given strain) the number of soluble expression hits was greater at 298 K compared with 310 K, while the number of hits was not increased by further lowering the temperature to 290 K (Fig. 3). This confirms the conventional wisdom that lower temperatures tend to be more effective, reflected in the protocols of several SPINE laboratories which routinely use a single lower temperature (*e.g.* 293 K used by Oxford and Strasbourg).

3.3.3. *E. coli* strains. Most groups screened for soluble expression in at least two *E. coli* strains, with six of the eight centres using a strain that co-expressed rare codons. The sparse-matrix screen carried out by the Marseille group comprised 12 different culture conditions for each of the 96

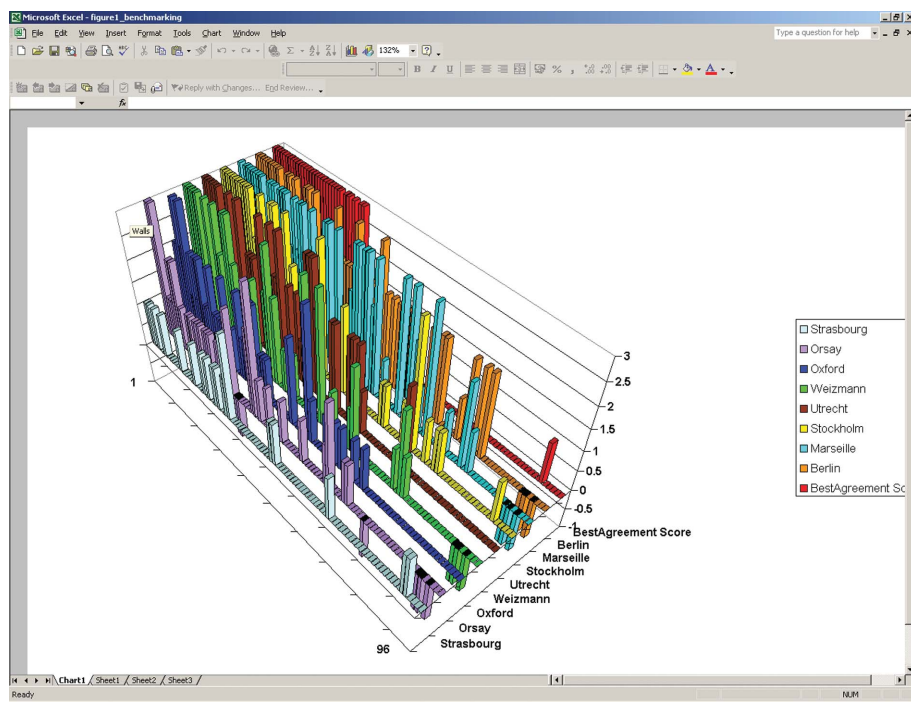


Figure 2 Plot of soluble expression results. The best soluble expression score for each target from each SPINE group ($BEST_{jk}$) is presented as a three-dimensional plot. The targets (x axis) have been ranked according to soluble expression score (y axis) from highest to lowest TOT_TARGET_k and the groups (z axis) ranked by laboratory score from highest to lowest TOT_LAB_k .

constructs, varying not only the post-induction temperature but also the *E. coli* strain (Vincentelli *et al.*, 2005). The results (Fig. 3) show the benefit of this approach. Any one strain/temperature combination gave, on average, a TOT_LAB score of 58 ($BEST > 0$ for 30 targets), whereas using four strains at three temperatures gave an aggregate TOT_LAB of 114 ($BEST > 0$ for 49 targets). The different strain/temperature combinations can be ranked with Rosetta (DE3)-pLysS/298 K and Rosetta (DE3)-pLysS/310 K performing best, giving 76 and 72 soluble proteins, respectively. This suggests that for this test set, containing mostly plasmids encoding eukaryotic proteins, the use of a codon-enhanced strain (*e.g.* strains carrying the pLysSRARE plasmid) has a greater positive contribution than the culture temperature. This is in line with the data from the Berlin group ($TOT_LAB_{Berlin} = 119$, $BEST > 0$ for 50 targets), where the Rosetta (DE3) strain was used at 293 K and the benefit of using such strains for the expression of eukaryotic proteins has been

reported previously (Sorensen & Mortensen, 2005). It is interesting to note, however, that the Marseille screening procedure led to a gain of 19 additional soluble expression hits from the full 12-condition screen.

3.3.4. Timing of induction. Induction is generally performed at early or mid-log phase; however, there are reports that induction in late-log phase (Galloway *et al.*, 2003) or even stationary phase (Ou *et al.*, 2004) can influence both total expression levels and solubility. To investigate this effect, the Utrecht group (Folkers *et al.*, 2004) performed induction at early, mid- and late-log phase and at early stationary phase corresponding to an OD₆₀₀ of 0.8, 1.6, 2.3 and 3.0, respectively. Overall, induction at early log phase gave the best results. Stationary phase induction was counterproductive (total expression was completely lost for half of the targets and decreased for the remainder), while soluble expression was reduced even more. In Utrecht 18 targets gave soluble expression; for these, soluble expression of eight was not influenced by more than a factor of two by the timing of induction, whilst for five early induction was beneficial and for five mid- or late-log gave the highest yields of soluble protein (Fig. 4). Thus, the timing of induction may be a useful parameter to vary in small-scale expression screens.

3.3.5. Autoinduction. The auto-induction protocol of Studier (2005) for pET-based T7 promoter vectors is amenable to high-throughput applications and a number of suitable media preparations are available commercially. The Oxford and Strasbourg protocols used auto-induction medium (alone in Strasbourg and alongside IPTG induction in Oxford), but the results from the two laboratories are surprisingly different (TOT_LAB scores of 18 and 44, respectively, and BEST > 0 scores of 17 and 24). This discrepancy could reflect the shorter growth time at lower temperature in Strasbourg (16 h at 293 K versus 24 h at 298 K in Oxford) or differences in strains and

media. However, the strains used [BL21 (DE3) and B834 (DE3)] are closely related and the results presented above suggest that the medium has relatively little effect on expression/solubility, so it seems likely that time and temperature account for much of the difference. The IPTG and auto-induction results of Oxford (the use of both induction methods is standard in that laboratory) were similar (SCALE = 12 and 14, respectively). Ten of these targets were common to both induction regimes, with four targets detected exclusively with auto-induction and two exclusively with IPTG-induction. However, in general the previous experience of both laboratories is that both IPTG and autoinduction usually give qualitatively similar results, in terms of target 'coverage', although autoinduction can result in higher levels of expression. Therefore, where soluble expression is detected using both methods, auto-induction provides a simpler procedure for scale-up, although the higher biomass obtained using auto-induction may affect subsequent sample processing.

3.3.6. Control of expression. Strasbourg observed leaky expression at the pre-culture stage for some targets that did not give expression after overnight culture using auto-inducible medium. This was probably a consequence of lactose contamination in the media (Studier, 2004), which can be eliminated by adding glucose to both pre-culture and culture media. Alternatively, the BL21-AI strain (bearing a chromosomal copy of the T7 RNA polymerase gene under the control of the arabinose-inducible *araBAD* promoter) provides a tighter control of T7-based plasmids. For this target set, there was no obvious advantage in using this strain since in the hands of the Weizmann group only one target that failed in the other groups gave soluble expression (see supplementary material).

3.3.7. Detection method. The results allow us to compare dot-blot (Stockholm and Marseille) and SDS-PAGE methods for the detection of soluble expression. As shown in Figs. 1 and 2, no disagreement was found for the insoluble and highly expressed constructs, but for the weakly expressed and/or

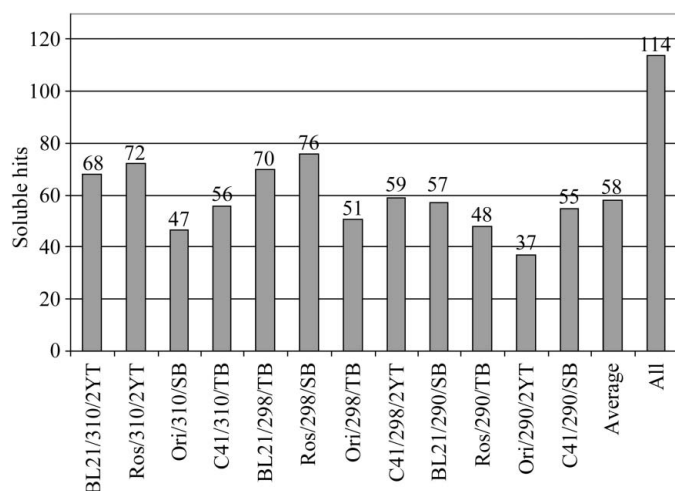


Figure 3

Effect of *E. coli* strains, temperature at induction and media on protein solubility. Summary of the results obtained by Marseille. The results of the sparse-matrix screen for soluble expression of the benchmark vectors are shown. The solubility level scores (0, 1, 2, 3) have been summed to give a total value for each strain/post-induction growth temperature/media combination and are indicated at the top of each bar.

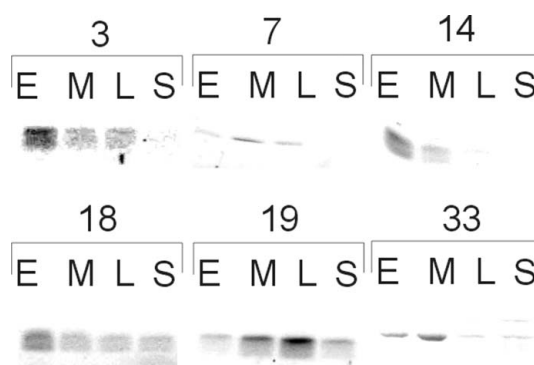


Figure 4

Influence of culture conditions on soluble protein production: OD₆₀₀. Plasmids were transformed in RosettaII (DE3) strain and induced during early (E), mid (M) or late (L) logarithmic growth phase or during early stationary phase (S) using 1 mM IPTG for 14 h at 293 K. Extracts prepared with PopCulture were affinity-purified using MagnaHis beads and analyzed by SDS-PAGE. Examples are shown where the OD₆₀₀ value influences soluble expression.

soluble there were large differences. The Stockholm dot-blot protocol used comparable amounts of cell lysate to the SDS-PAGE protocol of Oxford and Utrecht, but detected fewer expressed clones. In contrast, Marseille, by loading five times more purified protein, detected more soluble expressed protein from the panel of clones; indeed, not every clone with a positive dot-blot signal gave a detectable signal on SDS-PAGE. It therefore seems that by loading more protein it is possible to render the dot-blot method as sensitive as the routine SDS-PAGE method. Dot-blots are eminently suitable for automation; however, the lack of information about size and purity is a serious limitation. For instance, the Stockholm group picked up an erroneous, albeit weak, signal for one of three known negatives in this test using a dot-blot screen. Dot-blots are therefore best suited for initial screening in situations where the success rate is expected to be very low (such as the massively parallel analysis of constructs of proteins which are very difficult to express).

4. Conclusions

During the course of the SPINE contract, a number of partner laboratories have established rapid and cost-effective small-scale screening for soluble expression in *E. coli*. In conjunction with the parallel developments in upstream and downstream technologies, these have fundamentally changed the sample-preparation stages of structural biology. Our analysis of the SPINE expression-screening pipelines demonstrates that the different methodologies identify similar groups of best and worst expressing proteins which can give, in most cases, a good prediction of levels of soluble expression potentially attainable on scale-up. This is important since it enables effort downstream of cloning and expression screening to be invested in the most tractable targets. This study also highlighted the variability between protocols to detect proteins that fell between the two extreme scores, demonstrating that, at least for the anonymous often problematic targets chosen for this study (mainly eukaryotic and viral proteins), there is a substantial cohort for which the parameters chosen for the screening have a major effect on the expression. This suggests that the expression experiment has features in common with other crystallographic activities such as crystallization, which are conducted in a multi-parameter space. The present results do not allow us to dissect fully the correlations that may exist between the parameters in 'expression space', mainly because the difficult nature of targets led to a rather low success rate. Nevertheless, the effects of several parameters can be clearly discerned (as discussed above) and we would suggest that the 'sparse-matrix' approach of Marseille, whilst expensive in terms of the number of tests required per protein, could be augmented and refined in the light of these findings. Further work would then be required to establish sets of guidelines for the most effective strategy for the production, for a particular protein, of sufficient material for structural studies. The components of this strategy would include (i) construct optimization (discussed in Alzari *et al.*, 2006, with library based-methods to scan for expressible domains), (ii) the use of

homologous proteins from different species (see Siebold *et al.*, 2005), and (iii) the exploration of expression space, including the optimized prokaryotic screening discussed here, but also, especially if the proteins are particularly complex or subject to post-translational modifications, screening in eukaryotic expression systems (see, for example, Aricescu *et al.*, 2006).

This work was supported by the European Commission as part of SPINE, Structural Proteomics In Europe, contract No. QLG2-CT-2002-00988, under the Integrated Programme 'Quality of Life and Management of Living Resources' and by the German Federal Ministry of Education and Research (BMBF) through the National Genome Network, grant ID 01GR0472 to the Berlin Center.

References

- Abergel, C., Coutard, B., Byrne, D., Chenivresse, S., Claude, J. B., Deregnaucourt, C., Fricaux, T., Giancesini-Boutreux, C., Jeudy, S., Lebrun, R., Maza, C., Notredame, C., Poirot, O., Suhre, K., Varagnol, M. & Claverie, J. M. (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.
- Alzari, P. M. *et al.* (2006). *Acta Cryst.* **D62**, 1103–1113.
- Aricescu, A. R. *et al.* (2006). *Acta Cryst.* **D62**, 1114–1124.
- Busso, D., Poussin-Courmontagne, P., Rosé, D., Ripp, R., Litt, A., Thierry, J.-C. & Moras, D. (2005). *J. Struct. Funct. Genomics*, **6**, 81–88.
- Cornvik, T., Dahlroth, S. L., Magnusdottir, A., Herman, M. D., Knaust, R., Ekberg, M. & Nordlund, P. (2005). *Nature Methods*, **2**, 507–509.
- Dumon-Seignovert, L., Cariot, G. & Vuillard, L. (2004). *Protein Expr. Purif.* **37**, 203–206.
- Folkers, G. E., van Buuren, B. N. & Kaptein, R. (2004). *J. Struct. Funct. Genomics*, **5**, 119–131.
- Galloway, C. A., Sowden, M. P. & Smith, H. C. (2003). *Biotechniques*, **34**, 524–530.
- Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557–580.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Inoue, H., Nojima, H. & Okayama, H. (1990). *Gene*, **96**, 23–28.
- Knaust, R. K. & Nordlund, P. (2001). *Anal. Biochem.* **297**, 79–85.
- Nakata, Y., Tang, X. & Yokoyama, K. K. (1997). *Methods Mol. Biol.* **69**, 129–137.
- Ou, J., Wang, L., Ding, X., Du, J., Zhang, Y., Chen, H. & Xu, A. (2004). *Biochem. Biophys. Res. Commun.* **314**, 174–180.
- Scheich, C., Sievert, V. & Büssov, K. (2003). *BMC Biotechnol.* **3**, 12.
- Siebold, C., Berrow, N., Walter, T. S., Harlos, K., Owens, R. J., Stuart, D. I., Terman, J. R., Kolodkin, A. L., Pasterkamp, R. J. & Jones, E. Y. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 16836–16841.
- Sorensen, H. P. & Mortensen, K. K. (2005). *Microb. Cell Fact.* **4**, 1.
- Studier, F. W. (2004). Personal communication.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.
- Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* **D58**, 2109–2115.
- 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., Offant, J., Cambillau, C. & Bignon, C. (2005). *Anal. Biochem.* **346**, 77–84.
- Walter, T. S., Diprose, J. M., Mayo, C. J., Siebold, C., Pickford, M. G., Carter, L., Sutton, G. C., Berrow, N. S., Brown, J., Berry, I. M., Stewart-Jones, G. B., Grimes, J. M., Stammers, D. K., Esnouf, R. M., Jones, E. Y., Owens, R. J., Stuart, D. I. & Harlos, K. (2005). *Acta Cryst.* **D61**, 651–657.