

A. R. Aricescu,^a R. Assenberg,^a
R. M. Bill,^{b,‡} D. Busso,^c V. T. Chang,^d
S. J. Davis,^d A. Dubrovsky,^e
L. Gustafsson,^b K. Hedfalk,^b
U. Heinemann,^f I. M. Jones,^g
D. Ksiazek,^h C. Lang,ⁱ K. Maskos,^h
A. Messerschmidt,^h S. Macieira,^h
Y. Peleg,^j A. Perrakis,^k
A. Poterszman,^c G. Schneider,^d
T. K. Sixma,^k J. L. Sussman,^j
G. Sutton,^a N. Tarboureich,^l
T. Zeev-Ben-Mordehai^j and
E. Yvonne Jones^{a,*}

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

^bDepartment of Chemistry and Biosciences, Chalmers University of Technology, PO Box 462, SE-40 530 Göteborg, Sweden, ^cInstitut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent

Fries, BP 163, 67404 Illkirch CEDEX, France,

^dWeatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, England,

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

^fMax-Delbrück-Center for Molecular Medicine, Department of Crystallography, Robert-Rössle-Strasse 10, D-13125 Berlin, Germany, ^gSchool of Biological Sciences, University of Reading, Whiteknights,

PO Box 217, Reading RG6 6AH, England, ^hMax-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Am Klopferspitz 18,

82152 Martinsried, Germany, ⁱBiotechnology Institute, Technical University of Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany, ^jThe Israel Structural

Proteomics Centre, Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel,

^kDivision of Molecular Carcinogenesis, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands, and ^lEMBL-Grenoble, c/o ILL, BP 181, 6 Rue Jules Horowitz, F-38042 Grenoble CEDEX 9, France

[‡] Present address: School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, England.

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

Eukaryotic expression: developments for structural proteomics

The production of sufficient quantities of protein is an essential prelude to a structure determination, but for many viral and human proteins this cannot be achieved using prokaryotic expression systems. Groups in the Structural Proteomics In Europe (SPINE) consortium have developed and implemented high-throughput (HTP) methodologies for cloning, expression screening and protein production in eukaryotic systems. Studies focused on three systems: yeast (*Pichia pastoris* and *Saccharomyces cerevisiae*), baculovirus-infected insect cells and transient expression in mammalian cells. Suitable vectors for HTP cloning are described and results from their use in expression screening and protein-production pipelines are reported. Strategies for co-expression, selenomethionine labelling (in all three eukaryotic systems) and control of glycosylation (for secreted proteins in mammalian cells) are assessed.

Received 24 March 2006

Accepted 31 July 2006

1. Introduction

Target-protein expression presents one of the first hurdles to overcome in a structure determination. The Structural Proteomics In Europe (SPINE) consortium (<http://www.spineurope.org>) is committed to working predominantly on high-value (in terms of impact on human health) viral and human targets despite the observation that many such proteins are notoriously intractable targets for expression in *Escherichia coli*. In prokaryotes, the lack of post-translation modification, limited disulfide-bond formation and the absence of various chaperones often hinder the generation of properly folded fully functional eukaryotic proteins. A variety of eukaryotic expression systems have been developed in response to such problems. Although expression in these eukaryotic systems is, in general, time-consuming and more expensive than expression in prokaryotic systems, many structural biologists studying viral and human proteins have found that they often provide the only route forward. This was recognized at the earliest planning stages of SPINE, with the acceptance that in order to successfully tackle targets from a broad range of protein families, high-throughput (HTP) methodologies for eukaryotic expression would be required.

The development and implementation of HTP eukaryotic expression methodologies constituted SPINE workpackage 2. At the time of SPINE's inception (2002), the objectives of this workpackage were an essentially novel aspect of the European enterprise in HTP structural biology, as the structural genomics pipelines being developed and tested in the US and Japan were largely based on prokaryotic or cell-free expression systems and primarily targeted bacterial proteins

(Stevens, 2004). The challenge for SPINE was therefore twofold: *ab initio* development of robust HTP methodologies for eukaryotic expression in a subset of partner laboratories, followed by dissemination of these technologies to laboratories with little or no previous experience of such expression systems.

Eukaryotic expression systems can largely be grouped into three categories based on the nature of the cellular system used; namely, yeast, insect cells (the basis for baculovirus expression) and mammalian cells. At the start of SPINE, each of the available eukaryotic expression systems had obvious strengths, but also perceived weaknesses which hindered their more widespread use in standard structural biology laboratories and presented obstacles to their application in a HTP *modus operandi*.

Yeasts are single-cell eukaryotic hosts which combine some of the advantages of prokaryotic and eukaryotic based expression systems; for example, they are physically robust and amenable to high-density fermentation but possess the necessary cellular machinery to carry out post-translational modifications. The methylotrophic yeast *Pichia pastoris* gives high yields of recombinant proteins (Cereghino & Cregg, 2000), can be grown to high cell densities using defined minimal media and offers a cost-effective method for ¹³C-labelled protein production for NMR-based structural analyses (Laroche *et al.*, 1994). Typically, genes of interest are expressed under the control of the strong and tightly regulated *P. pastoris* alcohol oxidase 1 (AOX1) promoter. Baker's yeast, *Saccharomyces cerevisiae*, provides an alternative to *P. pastoris*, but with genes of interest expressed under the control of a different promoter; for example, the copper inducible metallothionein (CUP1) promoter.

Baculovirus expression of recombinant proteins in insect cells had, over the two decades before the start of SPINE, become a well established method for many proteins that are difficult to express in *E. coli* (Smith *et al.*, 1983; Kost *et al.*, 2005), during which time technological advances had increased its potential as a HTP methodology (Albala *et al.*, 2000). Over the lifetime of SPINE, earlier developments designed to improve the methodology of recombinant virus isolation, including positive selection of recombinant plaques (*e.g.* Vialard *et al.*, 1990), improved recovery of recombinants (Kitts & Possee, 1993) and the development of baculovirus recombination in yeast and *E. coli* (Patel *et al.*, 1992; Luckow *et al.*, 1993), the latter commercialized as the Bac-to-Bac system (Invitrogen), have given way to advances designed specifically for HTP use. For example, an alternate method of recombinant isolation has been developed by Invitrogen (BaculoDirect) to integrate baculovirus expression systems into its proprietary Gateway cloning system. In BaculoDirect *in vitro* (Gateway-based) recombination occurs between a suitable destination vector carrying the gene of interest and a baculovirus genome carrying a pseudo-lethal gene which is swapped for the gene of interest through the clonase recombination process. The recombinant virus can then be directly transfected into insect cells, where drug selection is applied to counter-select the parental virus.

Protein production using mammalian cell-based expression systems has not been widely used by structural biologists, but pre-SPINE it had already proved very effective in a number of cases, particularly for the production of secreted proteins. For example, stable expression in Chinese hamster ovary CHO cells (Cockett *et al.*, 1990) had been used successfully to produce proteins for a number of structure determinations (for example, Jones *et al.*, 1992; Casasnovas *et al.*, 1997; Wu *et al.*, 1997), but was generally perceived to be a specialist and expensive methodology requiring significant expertise in and facilities for tissue culture. In addition, the time scales are long, typically one to two months, for selection of stable clones expressing the protein of interest at sufficiently high levels. However, the development and streamlining of protocols for the transient expression of proteins in mammalian cells, such as human embryonic kidney (HEK) 293 cells (Meissner *et al.*, 2001; Durocher *et al.*, 2002; Aricescu, Lu *et al.*, 2006), now offers a methodology that is potentially compatible with HTP approaches.

We summarize here results and conclusions drawn from studies carried out in a subset of SPINE laboratories to assess the applicability of various eukaryotic expression methods for HTP structural biology. In line with the philosophy of parallelization and miniaturization underlying SPINE HTP strategies, emphasis was placed on the development of systems and protocols to facilitate rapid and efficient testing of multiple constructs in a variety of organisms/strains. Robust protocols for selenomethionine (SeMet) labelling and, for secreted proteins, methods to control the extent and heterogeneity of glycosylation are of particular importance for the use of eukaryotic expression systems in structural biology and are specifically addressed by developments and results from SPINE laboratories.

2. Materials and methods

To date, more than half of the laboratories in the SPINE consortium have tested eukaryotic expression systems for the production of particular targets (often through SPINE-based collaborations), but only a limited subset have used such systems on a regular basis. These laboratories have developed semi-automated approaches for testing protein expression in eukaryotic systems to parallel the high-throughput (HTP) techniques they have implemented for *E. coli*-based expression. In the following three subsections, we survey the approaches taken in the SPINE laboratories to streamline protocols for the production of proteins in yeast, baculovirus (insect cell) and mammalian cell-based expression systems.

2.1. Yeast

Four SPINE laboratories have reported results from yeast-based expression systems. Of these one, Göteborg, has specialized in the optimization of large-scale fermentation methods for the production of particular high-value target proteins (for example, a spinach plasma membrane aquaporin; Törnroth-Horsefield *et al.*, 2005). By systematically quanti-

fying cultures in high-performance bioreactors under tightly defined growth regimes, the group has examined the reasons for successes and failures in recombinant membrane-protein production in yeast (Bonander *et al.*, 2005). Of the other three SPINE partners (Berlin, Munich and Weizmann) that have investigated the use of yeast-based expression systems, only Berlin has experience of running a significant number of targets though in a pipelined approach and the methods they have developed are reviewed below, followed by protocols for co-expression of proteins as implemented at the Weizmann.

2.1.1. HTP cloning and expression. The Berlin group has reported systems for intracellular and extracellular expression of human proteins in the yeasts *S. cerevisiae* and *P. pastoris*. HTP methods were introduced wherever possible, including parallel cloning and transformation, parallel micro-scale expression and standardized fermentation and purification. Vectors were constructed to enable easy shuttling of cDNA sequences between yeast and *E. coli* expression systems. Details of the micro-scale (96-well format) processes developed for HTP cloning and expression are based on published protocols for *S. cerevisiae* (Holz *et al.*, 2002) and *P. pastoris* (Boettner *et al.*, 2002). In brief, for both yeasts the vector design was such that expressed protein was produced with both an N-terminal His₆ tag and a C-terminal StrepII tag to facilitate subsequent purification by two-step affinity chromatography. Expression was regulated by the CUP1 promoter in *S. cerevisiae* and the AOX1 promoter in *P. pastoris*. The clones selected using the small-scale (1 ml) expression screening methods were then grown in bioreactors using protocols detailed in Holz *et al.* (2003) and Prinz *et al.* (2004).

Refinement of these methods focused on the *S. cerevisiae* system. Mutant strains were constructed to increase expression efficiency. The most important mutation used proved to be the pep4 mutant, which is devoid of the major yeast protease and shows a decreased activity of all other proteases. A methionine-auxotrophic mutant was constructed to allow the incorporation of SeMet in the expressed proteins using a feeding regime of low SeMet concentration in the logarithmic growth phase and high concentration during the induction phase (Turnbull *et al.*, 2005). Cultivation and expression strategies were established for optimal protein yield under scale-up conditions (2–51 fed-batch fermentation) and a three-step chromatography-based protocol (including Talon matrix, StrepTactin and a gel-filtration step) was developed to isolate the recombinant proteins to high purity.

To test this pipeline human cDNAs, which had previously been cloned in the *E. coli* vector pQStrep2, were subcloned in *S. cerevisiae* by recombination-based cloning. The first step in this strategy was to amplify the complete expression cassette from the *E. coli* vector containing the target cDNA by high-fidelity polymerase-mediated PCR using 'recombination primers'. Flanking recombination sequences (40 nucleotides each), which are homologous to the CUP1 promoter and to the terminator region of the yeast vector pYEXTHS-BN, respectively, were thus added to the 5' and 3' ends of the expression cassette. The PCR products were co-transformed with the linearized expression vector pYEXTHS-BN in yeast

and the expression cassette was integrated by homologous recombination. Correct integration was confirmed by analytical PCR and sequencing analysis. HTP expression screening and IMAC purification of the expressed fusion proteins was carried out in 96-well format. Two PCR-verified yeast clones of each cDNA insert were screened for protein expression. Clones were checked by Western blot analysis of their total cellular proteins by using the PentaHis antibody (Qiagen). Proteins were purified under native conditions from cleared cell lysates using the amino-terminally fused His₆ tag and the resulting eluates were assessed by using the C-terminal StrepII-tag to detect the proteins and to confirm the full-length translation of the gene products.

2.1.2. Co-expression. The Weizmann group has, like the Berlin group, included expression in yeast as part of a unified strategy for HTP structural proteomics (Albeck *et al.*, 2005). Although not implemented in HTP mode, they have also developed protocols for the co-expression of two proteins in *P. pastoris*, a sequential transformation procedure which requires a two-step selection process. Initially, the gene for target 1 was cloned into the *P. pastoris* expression vector pPIC9K (Invitrogen) with a removable N-terminal His₆ tag and then transformed into *P. pastoris* GS115 strain selecting for a complementation *his-4* mutation. Target gene 2 was cloned into the expression vector pPICZ α (Invitrogen) without a tag. Transformation was then performed into a selected yeast clone harbouring multi-copies of the gene for target 1. Selection for target 2 gene integration and multi-copy clone selection was performed using the antibiotic zeocin. Following expression, initial purification of the target 1–target 2 complex used Ni-NTA agarose beads.

2.2. Baculovirus

Baculovirus-infected insect cells are the most commonly used eukaryotic expression system in SPINE (six of the partner laboratories, Amsterdam, Grenoble, Munich, Oxford, Strasbourg and Weizmann, report using this system on a regular or semi-regular basis and one sub-contractor group, Reading, has performed extensive development work on it).

2.2.1. HTP cloning and expression. During the course of the SPINE project, the partner laboratories used a broad range of cloning strategies (Alzari *et al.*, 2006); however, the overall trend was to move away from ligation-dependent cloning and three groups (Oxford, Reading and Strasbourg) have reported significant development work to streamline baculovirus methodologies. A key, and traditionally cumbersome, step is the generation of the recombinant baculovirus. The Reading group described genetic modification of the baculovirus genome to ensure 100% recombinant formation (Zhao *et al.*, 2003). In brief, the strategy uses a defective baculovirus genome that is rescued through recombination with a co-transfected plasmid containing the gene of interest. A commercialized version of this methodology has been implemented in the Oxford laboratory (see below). The Reading group also piloted a combined approach to *E. coli* and insect-

Table 1
Vectors.

Vector name	Description	Originator
pOPBac1	pBac2 (Novagen) baculovirus transfer vector adapted for Gateway, incorporates N-His ₆ tag	Oxford
pOPBac2	pBac2 (Novagen) baculovirus transfer vector adapted for Gateway	Oxford
pOPBac3	pBac2 (Novagen) baculovirus transfer vector adapted for Gateway, incorporates a C-terminal Fc+His ₆ tag	Oxford
pOPINE	pTriEX 2 modified for In-Fusion cloning, incorporates either a N-His ₆ or C-His ₆ tag depending upon site of cloning	Oxford
pOPINF	pTriEX 2 modified for In-Fusion cloning, incorporates either a N-His ₆ followed by a 3C protease cleavage site or C-His ₆ tag depending upon site of cloning	Oxford
pOPING	pTriEX 2 modified for In-Fusion cloning, incorporates a signal sequence for secretion in mammalian/insect cells	Oxford
pLEXm	pCAGGS derivative containing the chicken β -actin promoter and hCMV enhancer	Oxford
pHLsec	pLEXm derivative with resident signal sequence and C-terminal His ₆ tag	Oxford
pHLsec-FcHis	pHLsec derivative with C-terminal 3C protease Fc-His ₆ tag.	Oxford
pTriEX-MBP-Sfi	pTriEX1.1Sfi derivative for N-terminal fusion with secreted MBP and C-terminal His ₆ tag	Reading
pAC3CFcHis	pBacpAc (Clontech) derivative modified for directional cloning <i>via</i> Sfi1 sites with C-terminal 3C protease Fc-His ₆ tag	Reading
pPIC3.5K-Dest1	A Gateway-compatible destination vector for expression of intracellular proteins in <i>P. pastoris</i> . Based on the pPIC3.5K vector (InVitrogen). Contains an N-His ₆ tag.	Weizmann
pPIC9K-Dest1	A Gateway-compatible destination vector for expression of extracellular proteins in <i>P. pastoris</i> . Based on the pPIC9K vector (InVitrogen). Contains an N-His ₆ tag.	Weizmann
pBacGGWH	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-GST and C-His ₆ tags	Strasbourg
pBac0GW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway for native protein	Strasbourg
pBacFGW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-Flag tag	Strasbourg
pBacHGW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-His ₆ tag	Strasbourg
pBacAGW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-haemagglutinin tag	Strasbourg
pBacRGW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-Strep tag	Strasbourg
pBacCGW	pFastBac-1 (InVitrogen) baculovirus transfer vector adapted for Gateway, incorporates N-CBP tag	Strasbourg
pAC8C	pBacpAC8 (Clontech) modified for <i>NdeI</i> – <i>Bam</i> HI cloning, incorporates either a N-CBP tag followed by a 3C protease cleavage site	Strasbourg
pAC8O	pBacpAC8 (Clontech) modified for <i>NdeI</i> – <i>Bam</i> HI cloning, incorporates either a N-protein A tag followed by a 3C protease cleavage site	Strasbourg
pAC8F	pBacpAC8 (Clontech) modified for <i>NdeI</i> – <i>Bam</i> HI cloning, incorporates either a N-Flag tag followed by a 3C protease cleavage site	Strasbourg
pAC8G	pBacpAC8 (Clontech) modified for <i>NdeI</i> – <i>Bam</i> HI cloning, incorporates either a N-GST tag followed by a 3C protease cleavage site	Strasbourg
pAC8X	pBacpAC8 (Clontech) modified for <i>NdeI</i> – <i>Bam</i> HI cloning, incorporates either a N-thioredoxin tag followed by a 3C protease cleavage site	Strasbourg

cell expression through the use of dual promoter vectors (Xu & Jones, 2004; Chambers *et al.*, 2004). This multi-promoter strategy was adopted in Oxford, where the pTriEX 2 vector was modified for In-Fusion cloning (see Table 1). Oxford also adapted the pBac2 (Novagen) baculovirus transfer vector to allow Gateway cloning. Similarly, Strasbourg developed a set of Gateway-based vectors (see Table 1) which, after recombination to create the baculovirus, encode N-terminal fusion(s) as well as a C-terminal His₆ tag in frame with the sub-cloned ORF. This design followed the same model as that used in Strasbourg for prokaryotic expression vectors (*i.e.* providing the possibility of inserting a new fusion encoding sequence using specific restriction sites located both upstream and downstream of the Gateway cassette; Busso *et al.*, 2005; D. Busso, in preparation).

The pipeline approaches used by SPINE laboratories for expression screening and protein production were broadly similar; the only major difference was at the stage of recombinant virus production. In Strasbourg, recombinant baculovirus DNA was generated in *E. coli* using the Bac-to-Bac system (Life Technologies). In Oxford, Gateway or In-Fusion ligation-independent cloning was used, either *via* the entry plasmid (pDONR) for Gateway or directly into a pTriEX-

derived vector (pOPINE or pOPINF; Table 1) for In-Fusion cloning. For Gateway, cloning targets were then transferred to a destination vector compatible with *in vivo* recombination, pOPBAC2 (Table 1). Co-transfection into Sf9 cells of the pOPBAC2 or pTriEX constructs together with FlashBac baculovirus DNA (Oxford Expression Technologies, UK) was then used to generate the initial virus stock. 5 d following transfection the virus supernatant was collected and used to infect Sf9 and TnHi5 cells at an estimated MOI of 1 for 3 d before expression analysis.

Many of the automated procedures developed for small-scale expression screening in *E. coli* can be applied to the baculovirus system. For example, the Strasbourg laboratory adapted the parallel culture in the deep-well blocks method described by Bahia *et al.* (2005) so that they use the same automated procedure (Berrow *et al.*, 2006) for screening both prokaryotic and eukaryotic expression. Briefly, cells were harvested by centrifugation, suspended in lysis buffer and disrupted using a 24-probe sonication head, after which expression and solubility were assessed by SDS-PAGE. Since all the constructs harbour a His₆ tag (either N- or C-terminal), automated mini-purification screening can be used as for prokaryotic expressed proteins. Soluble fractions are applied

onto affinity resin dispensed into a 96-deep-well culture plate. After extensive washing, bound proteins are analyzed on SDS-PAGE by adding directly loading buffer to the resin.

SPINE laboratories typically reported protein expression in flasks to be a convenient and adequate means of production for most protein targets; for example, in Strasbourg scale-up (1–2 l cultures) used Bellco flasks (several of which could be used in parallel for different targets). However, where larger scale production was required (for targets which gave low expression but were of high scientific value) Oxford and Reading established large-scale (5–10 l) suspension cultures of insect cells using disposable bioreactors (Wave Biotech).

The inclusion of His₆ tags to facilitate downstream protein purification is the favoured strategy of all the groups; however, since components in the insect-cell media interfere with binding of His tags to IMAC, the Oxford group modified a vector to encode a C-terminal rhinovirus 3C protease-cleavable Fc+His₆ tag to allow convenient protein A-based affinity purification of secreted products (Table 1).

2.2.2. SeMet labelling. As with the other eukaryotic expression hosts, the efficient incorporation of SeMet into the expressed proteins represents a potentially major block to any structure-determination pipeline based on expression in insect cells. The Oxford group investigated protocols for SeMet labelling in baculovirus-based insect-cell expression using two standard cell lines, Sf9 and High5 (Invitrogen), both grown in SF900II media. The cells were infected with wild-type baculovirus (AcMNPV) to produce polyhedra. 20 h post-infection, the media were removed, replaced with cysteine- and methionine-free SF900II media supplemented with dialysed FCS to 10% (v/v) and 150 mg l⁻¹ cysteine. After a further 4 h growth to deplete cellular methionine levels, SeMet was added to either 100 or 500 mg l⁻¹. Cells were evidently infected 72 h post-infection and were harvested. Polyhedra were purified as described in Hill *et al.* (1999) using centrifugation onto sucrose cushions, dissolved in carbonate buffer pH 10.5 and submitted for mass-spectroscopic analysis.

2.3. Mammalian cells

Three of the laboratories (Amsterdam, Munich and Oxford) have used transient expression in mammalian cells to produce target proteins. The cell lines used are all based on human embryonic kidney (HEK) 293 cells, which are adherent cells which are relatively robust, easy to culture and have a good growth rate (doubling in number approximately every day). HEK 293T (used in Amsterdam and Oxford) and HEK 293EBNA (used in Munich) are both HEK cell lines which have been immortalized. *N*-Acetylglucosaminyltransferase I-negative HEK 293S (HEK 293S GnT^I-) cells limit the N-linked glycosylation of expressed proteins (Reeves *et al.*, 2002; Chang *et al.*, manuscript in preparation) and have therefore been used for expression of secreted glycoproteins in Amsterdam and Oxford.

Details of protocols (including SeMet labelling) for transient expression in HEK 293T cells that are suitable for a standard structural biology laboratory are presented in

Aricescu, Lu *et al.* (2006). In Oxford these protocols were primarily applied to secreted protein targets. Briefly, DNA from an overnight bacterial culture was purified to an OD₂₆₀/OD₂₈₀ ratio of 1.8 or higher and used to transfect cells which had reached ~90% confluency. Polyethylenimine (PEI) was used as the transfection reagent at a DNA:PEI ratio of 1:1.5 and 3–4 d later conditioned media were ready for collection and protein purification. SeMet labelling was carried out by modifying the standard protocols, from transfection onwards, to use methionine-free Dulbecco's Modified Eagle's Medium (DMEM; MP Biomedicals) supplemented with L-glutamine, non-essential amino acids and 30 mg l⁻¹ SeMet. A series of mammalian expression vectors (pLEXm and modified versions thereof), designed for use in restriction-enzyme-based cloning are detailed in Aricescu, Lu *et al.* (2006) and the pTriEX 2 series of multi-promoter vectors, modified for In-Fusion cloning, are presented in Table 1.

3. Results and discussion

3.1. Development and use of vectors for protein production in eukaryotic systems

The strategy underlying vector development has been to facilitate efficient gene cloning into multiple vectors as well as different expression systems (prokaryotic as well as eukaryotic; Alzari *et al.*, 2006). A second unifying SPINE theme has been the incorporation of His₆ tags to allow standardized approaches to be developed for protein-expression screening and initial purification. As a result, vector development has been carried out for all three eukaryotic expression systems: yeast, baculovirus and mammalian.

3.1.1. Yeast. The Weizmann group has developed a set of Gateway-compatible vectors for internal and secreted protein expression in *P. pastoris*, both harbouring a removable N-terminal His₆ tag (Peleg *et al.*, unpublished work). Similarly, as detailed in §2.1.1, the Berlin group has constructed vectors for *S. cerevisiae* and *P. pastoris*, such that cDNA sequences can be easily shuttled between the yeast and *E. coli* expression systems, and has included coding for an N-terminal His₆ tag (plus a C-terminal StrepII tag) to facilitate purification. These vectors have been used routinely for HTP cloning and expression. For example, the Berlin group report that of 192 different cDNAs cloned in the yeast expression vector, 112 could be expressed as soluble proteins in *S. cerevisiae*, corresponding to a success rate of 58%. In total during the Protein Structure Factory project in Berlin (which in part pre-dated SPINE), several hundred recombinant yeast strains were established and as a result are available for protein purification. Typically, they have found the protein yield from a 1 l cultivation to be between 1 and 7 mg.

3.1.2. Baculovirus. The Reading group developed a vector-suite approach to HTP expression. This work also led to the description of a unified approach to baculovirus expression through the provision of both N-terminal or C-terminal fusion vectors (Zhao *et al.*, 2003; Xu & Jones, 2004). Using kinases as test proteins, the Reading group have shown that amino-

terminal fusion to maltose-binding protein (MBP) rescues expression of the poorly expressed human kinase Cot but has only a marginal effect on expression of a well expressed kinase IKK-2. MBP fusion was also shown to be a useful approach for several other kinases, including p21-activated kinase 4, SGK3, CDK9 and mitogen-activated protein kinase-activated protein kinase (MAPKAPK). In addition, the Reading group have demonstrated that tagging with green fluorescent protein provides convenient readout of expression and that fluorescence levels match the levels of protein observed by SDS-PAGE. Expression of protein using the same vectors *in vitro* showed that differences in yield were wholly dependent on the environment of the expressing cell and that the time of harvest and protease addition substantially affected the observed expression level for poorly expressed proteins, but not for well expressed proteins. Details of the pilot studies on rapid expression and data on the underlying basis of the expression level obtained are reported in Pengelley *et al.* (2006).

Similarly, in Strasbourg His₆ and GST tagging were systematically compared for several cancer-related targets including the XPD helicase, the glucocorticoid receptor and the CARM1 transcription factor. Several constructs designed to vary the domain boundaries were tested for each target. For these three proteins, expression of the constructs with an amino-terminal GST fusion provided the best results. In the case of the CARM1 protein, none of the 25 His₆-tag constructs that were tested led to the expression of a protein suitable for structural analysis, whereas four out of the 25 GST fusion proteins allowed the production of a soluble protein (Troffer-Charlier, in preparation). At the time of this report, crystals have been obtained for one of these constructs.

The Reading group have also developed a second set of vectors based on a similar cloning strategy but designed for the expression of secreted proteins with a number of tags including human Fc, TAP and His₆. Initial work with these vectors has centred on the expression of the Spike glycoprotein of SARS and has been reported by Yao *et al.* (2004). Some of the constructs described in that work were scaled up and workable quantities of protein (>1 mg) were obtained (Fig. 1). Unfortunately, crystallization trials using these proteins were unsuccessful and removal of the Fc tag was problematic. A number of other glycoproteins have since been cloned and expressed in a variety of tagged formats. These include coronavirus NL63 S1, influenza Vietnam H5, HIV gp120 outer domain and bovine viral diarrhoea virus (BVDV) E2 protein. These are types of proteins that are generally considered to be difficult targets and problems which have prevented the progress of these proteins into crystallization trials have included low expression levels, inability to remove the fusion partner efficiently and poor purification. However, of the above set of targets, the BVDV E2 protein (with C-terminal His₆ tag) has proved to be reproducibly purifiable at the 10 mg scale and has entered crystallization trials. The protein is a variant of the wild type in which one glycosylation site has been removed without loss of biological activity (measured as receptor binding) and has been described by Pande *et al.* (2005).

The strategies for producing proteins in insect cells are the most varied and complex of the three types of eukaryotic cells and timelines for three insect pipelines used in SPINE are shown in Fig. 2; namely, FlashBac and BaculoDirect (Oxford) and Bac-to-Bac (Strasbourg). The approaches differ by the technology used to generate the initial viral stock. For the Oxford systems, ligation-independent cloning (Gateway or In-Fusion) is followed by co-transfection of the target constructs together with FlashBac baculovirus DNA (Oxford Expression Technologies, UK) into Sf9 cells to obtain the initial virus stock. The BaculoDirect approach also requires an initial Gateway cloning step, but the cost per reaction is substantially higher and there is less flexibility in terms of construct design (*e.g.* addition of different fusion tags). The Bac-to-Bac approach (Strasbourg) is less expensive but more complex since it involves an additional *E. coli*-based step and takes around one week before recombinant virus is obtained. Bac-to-Bac provides a robust methodology for a semi-automated approach to baculovirus expression; however, it is somewhat slower than FlashBac and BaculoDirect, which also have the advantage that they can be readily automated, exemplified by Oxford/Brookes where cell seeding into 24-well plates, transfections, infections, viral dilutions and parallel expression screening have all been implemented on a simple liquid handling robot (King *et al.*, manuscript in preparation).

3.1.3. Mammalian. Transient expression in mammalian cells has initially been assessed in standard structural biology laboratory settings. For example, in Oxford the pLEXm vector and variants thereof have been used for expression tests of more than 40 constructs of extracellular proteins ranging widely in size (20–150 kDa) and topology. The results for a panel of 24 constructs (Aricescu, Lu *et al.*, 2006) indicate soluble expression (>1 mg l⁻¹) for 18 targets at levels of 1–40 mg l⁻¹. This methodology has proved sufficiently robust to yield crystal structures (for example, the MAM-Ig

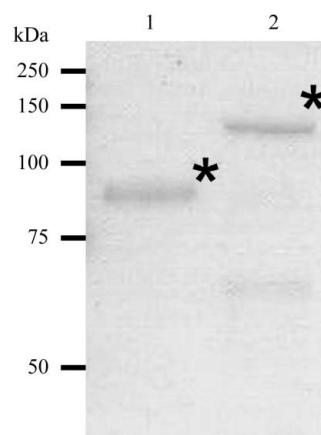


Figure 1
Purification of two fragments (*) of the SARS S protein as Fc fusions for crystal trial. The proteins were recovered from the supernatant of Sf9 cells 9 d post-infection and the recombinant proteins were captured and concentrated by lectin (*Lens culnaris*) chromatography. The lectin eluates were further purified by protein A affinity chromatography. The final yield was ~1 mg per litre of infected culture (10⁹ cells). The proteins shown are S1₁₉₋₄₁₀-Fc (lane 1) and S1₁₉₋₇₁₃-Fc (lane 2).

N-terminal domains of the receptor protein tyrosine phosphatase mu; Aricescu, Hon *et al.*, 2006) and is currently being adapted and optimized for automation in the Oxford HTP laboratory (the Oxford Protein Production Facility; N. Berrow & R. Owens, personal communication).

3.2. Co-expression

All three eukaryotic expression systems are amenable to co-expression of component proteins for *in vivo* formation of protein complexes.

The Weizmann group tested protocols for co-expression in *P. pastoris* (§2.1.2) using the extracellular domains of two *Drosophila* proteins, amalgam (Ama) and neurotactin (Nrt), involved in neuronal development, as targets 1 and 2, respectively. Both proteins were insoluble when expressed separately or co-expressed in *E. coli*. Following co-expression in yeast both proteins co-eluted from Ni-NTA agarose beads (Fig. 3). Since only the extracellular domain of Ama possesses a His₆ tag, this implies that the proteins were not only co-secreted but also formed a functional complex.

Albeck *et al.* (2006) report the experiences of the Amsterdam and Strasbourg groups for co-expression in baculovirus-infected insect cells. Four case studies are described of cytosolic complexes, for all of which expression of small quantities of soluble well behaved complex was achieved.

Oxford has assessed the efficacy of co-expression in transiently transfected mammalian cells for production of complexes between secreted proteins (Aricescu, Lu *et al.*, 2006). A co-transfection experiment for secreted components of a receptor–ligand complex yielded the complex with a significant improvement in expression levels over those observed on transfection of the individual components.

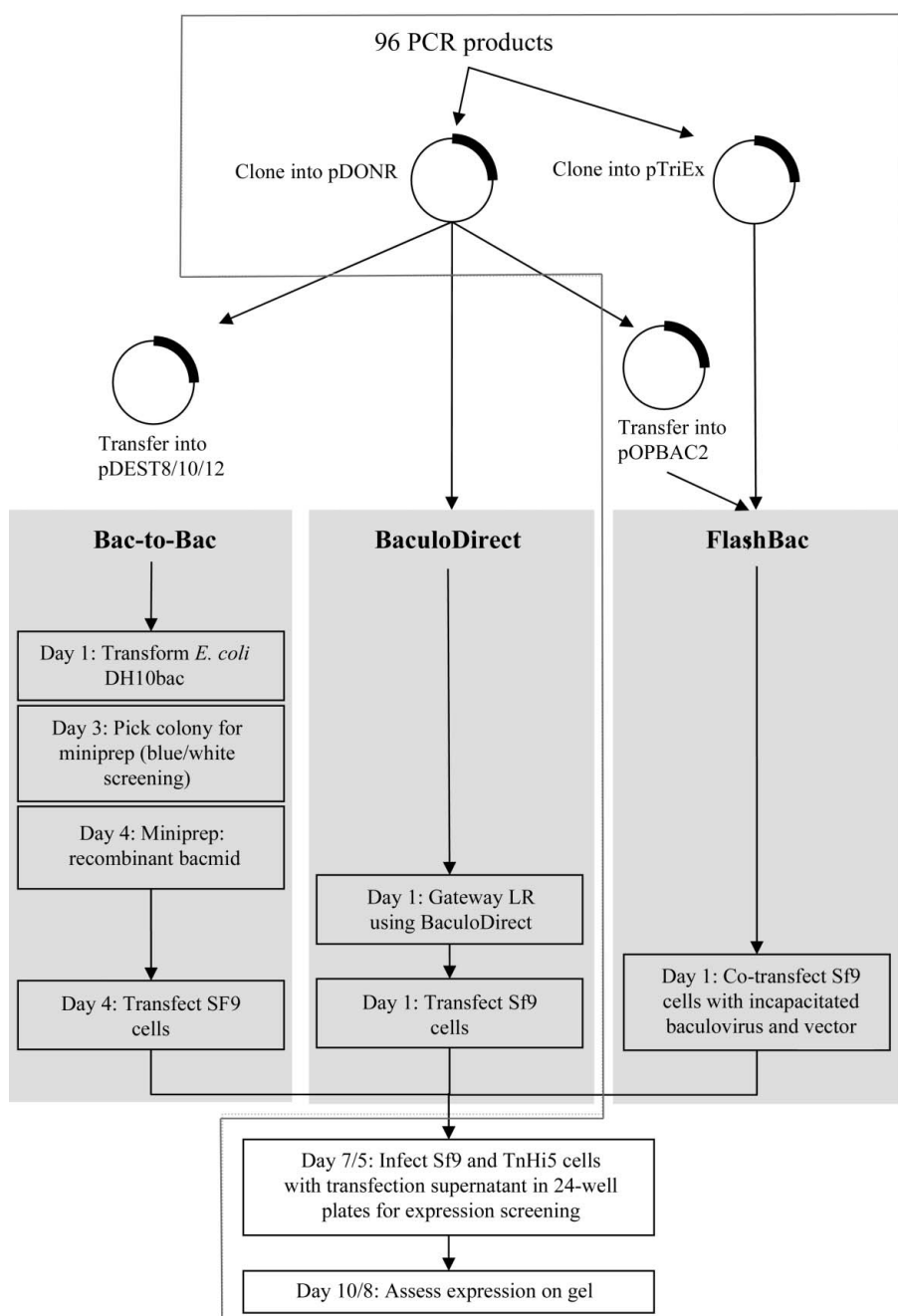


Figure 2 Pipeline approaches used by SPINE for baculovirus protein expression.

3.3. SeMet labelling

The ability to label proteins with SeMet is now generally considered to be a major requirement for any pipeline aiming to produce samples for protein crystallography. SPINE laboratories have investigated methods to meet this requirement in yeast, baculovirus and mammalian cell-based expression systems.

By using a methionine-auxotrophic mutant strain and an adapted feeding regime (see §2.1.1) the Berlin group achieved 40% SeMet incorporation in yeast, consistent with the levels documented in the literature (in the few examples reported pre-2004 none exceeded ~50% incorporation of SeMet; Bushnell *et al.*, 2001; Larsson *et al.*, 2002, 2003). However, one of the SPINE groups (Oxford), in a collaboration with the group of D. Bamford (University of Helsinki, Finland), have in a recent structure determination improved the efficacy of the protocol to achieve essentially complete (~98%)

SeMet labelling of a protein expressed in *S. cerevisiae* (Laurila *et al.*, 2005).

Although SeMet labelling has been reported for insect-cell expressed proteins (Bellizzi *et al.*, 1999; Carlson *et al.*, 2005), experience within the SPINE programme has suggested that extant protocols are not wholly reliable (Sutton *et al.*, unpublished observations). The Oxford group therefore carried out a series of experiments to refine previously published protocols for SeMet labelling in baculovirus-based insect-cell expression. Two standard cell lines were used, Sf9 and TnHi5 (Invitrogen), and both were grown in SF900II media with SeMet added to give concentrations of either 100 or 500 mg l⁻¹ (see §2.2.2). For each of the four experiments the level of SeMet incorporation was assessed using polyhedra produced in the cells after wild-type baculovirus (AcMNPV) infection. AcMNPV polyhedrin, the protein which forms polyhedra in the infected insect cells, contains six methionine residues. The incorporation levels for Sf9 cells were 1.03 and 3.14 Se atoms per protein for 100 and 500 mg l⁻¹ SeMet concentrations, respectively. For High5 cells the selenium incorporation rates were 2.11 and 3.78 per protein for 100 and 500 mg l⁻¹ SeMet concentrations, respectively. These initial results show two clear trends. SeMet incorporation is higher in High5 cells than Sf9 and the higher the concentration of SeMet in the media the greater the level of incorporation achieved (the maximum in this set of experiments being 63%).

To date, the Oxford group have used SeMet labelling for the structure determination of two secreted proteins transiently expressed in mammalian cells (Aricescu, Hon *et al.*, 2006; Aricescu *et al.*, manuscript in preparation). Approximately 60% SeMet incorporation was achieved (Aricescu, Lu *et al.*, 2006), similar to that reported above using the optimal protocol for baculovirus-based expression in insect cells; however, levels of protein expression were reduced. Despite the incomplete incorporation, the diffraction data collected

for the two SeMet-labelled proteins (at BM14, ESRF, Grenoble) were sufficient to phase the structures (in both cases there was approximately one Met residue per 100 amino acids).

3.4. The challenge of glycoproteins

The major bottlenecks in HTP structural biology pipelines which use bacterial expression are the production of soluble protein and of diffraction-quality crystals (DeLucas *et al.*, 2005). We have discussed how eukaryotic expression systems may provide a solution to the first of these problems for targets dependent on post-translational modifications. However, glycosylation may well stall the project at the second bottleneck since the flexible and/or heterogeneous glycans may hinder crystallization.

In order to surmount such problems pre-SPINE, the Oxford group relied on the stable expression of glycoproteins that are easily deglycosylated with endoglycosidase (EndoH), achieved by expressing the proteins in mutant Chinese hamster ovary (CHO) cell-derived *Lec3.2.8.1* cells (Davis *et al.*, 1993) or in wild-type CHO cells in the presence of the glucosidase I inhibitor *N*-butyldeoxynojirmycin (NB-DNJ; Davis *et al.*, 1995; Butters *et al.*, 1999). As discussed above, however, the selection and expansion of clones renders such methods incompatible with HTP. Within the SPINE framework, the Oxford laboratory has therefore explored the feasibility of extending these approaches to transient protein expression in mammalian hosts. Two strategies have been investigated: (i) converting the *Lec3.2.8.1* cell line into a host for transient expression and (ii) restricting N-glycan processing to oligomannose intermediates in other well established transient expression hosts, such as human embryonic kidney (HEK) 293T cells.

HEK293T cells have the advantage that they stably express the SV40 large T antigen, which drives the episomal replication of transiently transfected SV40 *ori*-containing plasmids, such as pEF-DEST51 (Heinzel *et al.*, 1988). Oxford therefore introduced the trans-activating SV40 and polyoma virus large T antigens into *Lec3.2.8.1* cells to enhance expression from pEF-DEST51 or the polyoma *ori*-containing vector, pSVE1-b1a (Heffernan & Dennis, 1991), respectively. Unexpectedly, co-transfection of SV40 large T-expressing plasmids with SV40 *ori*-containing plasmids (*i.e.* pEF-DEST51) diminished the already very weak transient expression of a test protein (*i.e.* 19A; Murphy *et al.*, 2002) in *Lec3.2.8.1* cells. Similarly, the stable expression of the polyoma large T antigen in *Lec3.2.8.1* cells failed to enhance transient expression from pSVE1-b1a.

Based on these observations, Oxford determined whether a suspension-adapted HEK293-derived cell line (293S/GnT1^{-/-}; Reeves *et al.*, 2002) lacking *N*-acetylglucosamine transferase 1 (GnT1) could be used to express readily deglycosylated protein. cDNA encoding the His₆-tagged extracellular region of the protein tyrosine phosphatase RPTPμ (Gebbinck *et al.*, 1991) which contains 12 glycosylation sites distributed over six domains, was cloned

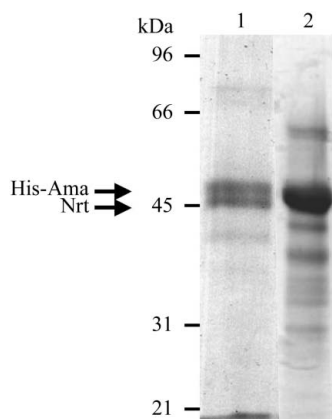


Figure 3

Co-expression of the His-Ama and Nrt proteins in *P. pastoris*. Cells harbouring both genes were induced for 2 d in BMMY medium. Proteins were analyzed on 12% SDS-PAGE followed by staining with GelCode (Pierce). Arrows indicate the predicted positions of the proteins. Lane 1, analysis of a 15 µl culture supernatant following 2 d induction; lane 2, proteins obtained upon elution from Ni-NTA agarose beads. Mass-spectrometric analysis revealed that the band at ~45 kDa (lane 2) contains peptides from both Ama and Nrt.

into the pLEXm expression vector, which was then transfected into 293S/GnT1^{-/-} cells. After 3 d, the protein was purified from the tissue-culture supernatant by metal-chelation chromatography. HPLC-based analysis of the released 2AB-labelled glycans indicated that whereas the large and heterogeneous N-glycans from the 293T cell line consist of multiantennary complex N-glycans typical of most mammalian expression systems, mutation of the GnT1 gene yields a pattern dominated by the Man₅GlcNAc₂ N-glycan. Moreover, virtually all the protein was sensitive to EndoH (Fig. 4). EndoH-treated RPTPμ formed crystals that diffract beyond 3 Å, whereas native glycosylated protein produced in 293T cells diffracted to >6 Å (Aricescu *et al.*, unpublished work), an observation consistent with the Oxford group's previous experience with this strategy of deglycosylation. Because yields from these cells are reduced by the absence of the SV40 large T antigen, however, Oxford have examined the effects of additional processing inhibitors on 293T cells, have attempted to derive ethyl methanesulfonate-mutated GnT1^{-/-}-deficient 293T cell lines (Chang *et al.*, in preparation) and have established methods for deglycosylating proteins expressed in insect-cell-based expression systems (Chang *et al.*, in preparation).

4. Use of eukaryotic expression: the SPINE experience

Prokaryotic expression is currently the pre-eminent tool for protein production in both standard structural biology and HTP-style laboratories. A survey of the relative usage in structural biology worldwide of prokaryotic and eukaryotic expression systems (based on Protein Data Bank depositions in 2004 and 2005) reveals that out of a total of nearly 7000 PDB entries deposited in the last 2 y, only 396 (less than 6%) record the use of an eukaryotic expression system. For these

396 entries the relative ratios for use of baculovirus, yeast and mammalian-based systems are approximately 3:2:1. These statistics are broadly representative of the level of eukaryotic expression system usage across most of the partner laboratories prior to the start of SPINE.

What lessons can be drawn from the SPINE experience of eukaryotic expression systems? Firstly, there are systems which are not considered promising candidates for use in HTP strategies. As noted above (§1 and §3.4) stable expression in mammalian (CHO) cells is a tried and tested route to protein production for structural studies (and has yielded a SPINE structure; Love *et al.*, 2003), but is not well suited to incorporation within a HTP-based strategy. Insect cells, like mammalian cells, can be used directly for stable expression of proteins. As part of SPINE, Stockholm tested a set of 25 human protein targets for stable expression in insect (S2) cells; however, the results were not encouraging since although ~50% of the targets were expressed as soluble proteins the levels of expression were in all cases less than 2 mg per litre of culture medium and in most cases were less than 1 mg per litre (G. Schneider; unpublished results). Thus, this strategy has not been pursued further and has not been detailed in the previous sections.

Within SPINE, baculovirus-infected insect cells have remained the most frequently used eukaryotic system; however, mammalian cells appear poised to overtake yeast as the second most used system. Several of the partners have implemented eukaryotic expression as a standard route for production of proteins that fail to give soluble expression in HTP *E. coli*-based expression screening; to date, this has been predominantly for the expression of human rather than pathogen protein targets (see Banci *et al.*, 2006; Fogg *et al.*, 2006). The success rates reported by SPINE laboratories for the soluble expression of human and viral target proteins in *E. coli*-based systems are 20–30% (see Alzari *et al.*, 2006); in comparison, insect and mammalian cell expression systems have delivered success rates of 45 and 76%, respectively (see Banci *et al.*, 2006). Whilst these success rates for the eukaryotic expression systems are still based on a relatively small sample set of SPINE targets (which is biased in terms of certain protein families, *e.g.* kinases, nuclear receptors, secreted proteins), they have clearly provided valuable rescue routes for high-value SPINE targets. The results for yeast-based expression are complicated by the small number of specifically SPINE target constructs tested (18 in total reported by the Amsterdam, Munich and Weizmann laboratories); for this small sample the success rate, 22%, was similar to that for *E. coli*. However, one of the SPINE laboratories, Berlin, has run a significant number of human proteins through a yeast-based expression pipeline and reports a success rate of 58% for soluble expression (§1.1), which is double that obtained in *E. coli*.

The commitment of the SPINE Partners to work predominantly on high value (in terms of impact on human health) but potentially difficult viral and human targets has demanded truly *ab initio* development of HTP methodologies for eukaryotic expression. In general, the implementation of

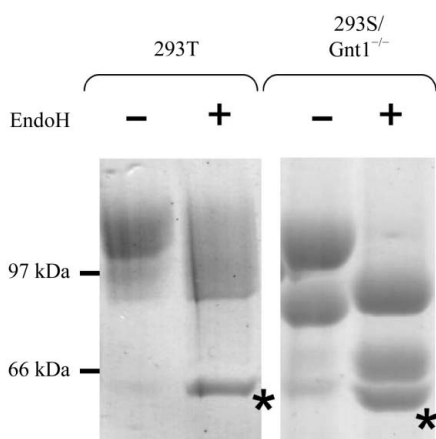


Figure 4 Deglycosylation of the receptor tyrosine phosphatase RPTPμ expressed transiently in 293T and 293S/GnT1^{-/-} cells. 5 mg of purified protein was treated with 250 U of endoglycosidase (EndoH) at pH 5.2 for 6 h at 310 K in each case. The samples were then analysed by SDS-PAGE under reducing conditions; the band marked with an asterisk is EndoH. Expression of RPTPμ in 293S/GnT1^{-/-} cells leads to a larger fraction of the protein being 'nicked'. In contrast to the partial EndoH-sensitivity of the 293T-derived material, the 293S/GnT1^{-/-}-derived protein is completely EndoH-sensitive.

yeast-based expression pipelines within SPINE has been limited and is currently not the favoured option for the majority of the groups, whereas baculovirus has delivered the most consistent success rates across the consortium. In addition to work within SPINE laboratories, much progress has been made elsewhere in the development of the baculovirus system; the use of unified vectors and robotics (Albala *et al.*, 2000), transfection in suspension and deep-well culture of insect cells (Bahia *et al.*, 2005; McCall *et al.*, 2005) and streamlining the overall process of recombinant baculovirus isolation (Phillips *et al.*, 2005) have all contributed to HTP baculovirus expression such that its systematic use, for example for herpesvirus open-reading-frame-encoded proteins, has been described (Gao *et al.*, 2005). Even the most streamlined system for expression screening in baculovirus takes approximately one week longer than a system based on transient expression in mammalian cells. Mammalian cell-based expression has, over the course of SPINE, emerged as a fast, robust and cost-effective method for efficient small-scale expression screening. For large-scale protein production comparative studies (Oxford) on the performance of baculovirus and mammalian cell-based expression systems are in agreement with the commonly held view that yields of cytosolic proteins are typically higher in the baculovirus system. However, for secreted proteins the converse is observed; transient mammalian expression significantly outperforms baculovirus-based insect-cell expression. Thus, mammalian cell-based expression strategies appear poised to complement insect cell based approaches for HTP protein expression.

This work was funded by the European Commission as SPINE, contract No. QLG2-CT-20020-00988, under the Integrated Programme 'Quality of Life and Management of Living Resources' and by the Wallenberg Consortium North.

References

- Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., Davies, A. H., Lennon, G. G. & Clark, R. (2000). *J. Cell. Biochem.* **80**, 187–191.
- Albeck, S. *et al.* (2006). *Acta Cryst.* **D62**, 1184–1195.
- Albeck, S., Burstein, Y., Dym, O., Jacobovitch, Y., Levi, N., Meged, R., Michael, Y., Peleg, Y., Prilusky, J., Schreiber, G., Silman, I., Unger, T. & Sussman, J. L. (2005). *Acta Cryst.* **D61**, 1364–1372.
- Alzari, P. M. *et al.* (2006). *Acta Cryst.* **D62**, 1103–1113.
- Aricescu, A. R., Hon, W.-C., Siebold, C., Lu, W., van der Merwe, P. A. & Jones, E. Y. (2006). *EMBO J.* **25**, 710–712.
- Aricescu, A. R., Lu, W. & Jones, E. Y. (2006). *Acta Cryst.* **D62**, 1243–1250.
- Bahia, D. B., Cheung, R., Buchs, M., Geisse, S. & Hunt, I. (2005). *Protein Expr. Purif.* **39**, 61–70.
- Banci, L. *et al.* (2006). *Acta Cryst.* **D62**, 1208–1217.
- Bellizzi, J. J., Widom, J., Kemp, C. W. & Clardy, J. (1999). *Structure*, **7**, R263–R267.
- Berrow, N. S. *et al.* (2006). *Acta Cryst.* **D62**, 1218–1226.
- Boettner, M., Prinz, B., Holz, C., Stahl, U. & Lang, C. (2002). *J. Biotechnol.* **99**, 51–62.
- Bonander, N., Hedfalk, K., Larsson, C., Mostad, P., Chang, C., Gustafsson, L. & Bill, R. M. (2005). *Protein Sci.* **14**, 1729–1740.
- Bushnell, D. A., Cramer, P. & Kornberg, R. D. (2001). *Structure*, **9**, 11–14.
- Busso, D., Poussin-Courmontagne, P., Rose, D., Ripp, R., Litt, A., Thierry, J.-C. & Moras, D. (2005). *J. Struct. Funct. Genomics*, **6**, 81–88.
- Butters, T. D., Sparks, L. M., Harlos, K., Ikemizu, S., Stuart, D. I., Jones, E. Y. & Davis, S. J. (1999). *Proteins*, **8**, 1696–1701.
- Carlson, C. B., Bernstein, D. A., Annis, D. S., Misenheimer, T. M., Hannah, B. L., Mosher, D. F. & Keck, J. L. (2005). *Nature Struct. Mol. Biol.* **12**, 910–914.
- Casasnovas, J. M., Springer, T. A., Liu, J. H., Harrison, S. C. & Wang, J. H. (1997). *Nature (London)*, **387**, 312–315.
- Cereghino, J. L. & Cregg, J. M. (2000). *FEMS Microbiol. Rev.* **24**, 45–66.
- Chambers, S. P., Austen, D. A., Fulghum, J. R. & Kim, W. M. (2004). *Protein Expr. Purif.* **36**, 40–47.
- Cockett, M. I., Bebbington, C. R. & Yarranton, G. T. (1990). *Biotechnology*, **8**, 662–667.
- Davis, S. J., Davies, E. A., Barclay, A. N., Daenke, S., Bodian, D. L., Jones, E. Y., Stuart, D. I., Butters, T. D., Dwek, R. A. & van der Merwe, P. A. (1995). *J. Biol. Chem.* **270**, 369–375.
- Davis, S. J., Puklavec, M. J., Ashford, D. A., Harlos, K., Jones, E. Y., Stuart, D. I. & Williams, A. F. (1993). *Protein Eng.* **6**, 229–232.
- DeLucas, L. J., Hamrick, D., Cosenza, L., Nagy, L., McCombs, D., Bray, T., Chait, A., Stoops, B., Belgovskiy, A., Wilson, W., Parham, M. & Chernov, N. (2005). *Prog. Biophys. Mol. Biol.* **88**, 285–309.
- Durocher, Y., Perret, S. & Kamen, A. (2002). *Nucleic Acids Res.* **30**, E9.
- Fogg, M. J. *et al.* (2006). *Acta Cryst.* **D62**, 1196–1207.
- Gao, M., Brufatto, N., Chen, T., Murley, L. L., Thalakada, R., Domagala, M., Beattie, B., Mamelak, D., Athanasopoulos, V., Johnson, D., McFadden, G., Burks, C. & Frappier, L. (2005). *J. Proteome Res.* **4**, 2225–2235.
- Gebbink, M. F., van Etten, I., Hateboer, G., Suijkerbuijk, R., Beijersbergen, R. L., Geurts van Kessel, A. & Moolenaar, W. H. (1991). *FEBS Lett.* **290**, 123–130.
- Heffernan, M. & Dennis, J. W. (1991). *Nucleic Acids Res.* **19**, 85–92.
- Heinzel, S. S., Krysan, P. J., Calos, M. P. & DuBridge, R. B. (1988). *J. Virol.* **62**, 3738–3746.
- Hill, C. L., Booth, T. F., Prasad, B. V., Grimes, J. M., Mertens, P. P., Sutton, G. C. & Stuart, D. I. (1999). *Nature Struct. Biol.* **6**, 565–568.
- Holz, C., Hesse, O., Bolotina, N., Stahl, U. & Lang, C. (2002). *Protein Expr. Purif.* **25**, 372–378.
- Holz, C., Prinz, B., Bolotina, N., Sievert, V., Bussow, K., Simon, B., Stahl, U. & Lang, C. (2003). *J. Struct. Funct. Genomics*, **4**, 97–108.
- Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). *Nature (London)*, **360**, 232–239.
- Kitts, P. A. & Possee, R. D. (1993). *Biotechniques*, **14**, 810–817.
- Kost, T. A., Condreay, J. P. & Jarvis, D. L. (2005). *Nature Biotechnol.* **23**, 567–575.
- Laroche, Y., Storme, V., De Meutter, J., Messens, J. & Lauwereys, M. (1994). *Biotechnology*, **12**, 1119–1124.
- Larsson, A. M., Andersson, R., Stahlberg, J., Kenne, L. & Jones, T. A. (2003). *Structure*, **11**, 1111–1121.
- Larsson, A. M., Stahlberg, J. & Jones, T. A. (2002). *Acta Cryst.* **D58**, 346–348.
- Laurila, M. R. L., Salgado, P. S., Makeyev, E. V., Nettleship, J., Stuart, D. I., Grimes, J. M. & Bamford, D. H. (2005). *J. Struct. Biol.* **149**, 111–115.
- Love, C. A., Harlos, K., Mavaddat, N., Davis, S. J., Stuart, D. I., Jones, E. Y. & Esnouf, R. M. (2003). *Nature Struct. Biol.* **10**, 843–848.
- Luckow, V. A., Lee, S. C., Barry, G. F. & Olins, P. O. (1993). *J. Virol.* **67**, 4566–4579.
- McCall, E. J., Danielsson, A., Buchs, M., Geisse, S. & Hunt, S. (2005). *Protein Expr. Purif.* **42**, 29–36.
- Meissner, P., Pick, H., Kulangara, A., Chatellard, P., Friedrich, K. & Wurm, F. M. (2001). *Biotechnol. Bioeng.* **75**, 197–203.

- Murphy, J. J., Hobby, P., Vilarino-Varela, J., Bishop, B., Iordanidou, P., Sutton, B. J. & Norton, J. D. (2002). *Biochem. J.* **361**, 431–436.
- Pande, A., Carr, B. V., Wong, S. Y. C., Dalton, K., Jones, I. M., McCauley, J. W. & Charleston, B. (2005). *Virus Res.* **114**, 54–62.
- Patel, G., Nasmyth, K. & Jones, N. (1992). *Nucleic Acids Res.* **20**, 97–104.
- Pengelley, S. C., Chapman, D. A., Abbott, W. M., Lin, H. H., Huang, W. & Dalton, K. (2006). *Protein Expr. Purif.* **48**, 173–181.
- Phillips, B., Rotmann, D., Wicki, M., Lorenz, M., Mayr, L. M. & Forstner, M. (2005). *Protein Expr. Purif.* **42**, 211–218.
- Prinz, B., Schultchen, J., Rydzewski, R., Holz, C., Boettner, M., Stahl, U. & Lang, C. (2004). *J. Struct. Funct. Genomics*, **5**, 29–44.
- Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 13419–13424.
- Smith, G. E., Summers, M. D. & Fraser, M. J. (1983). *Mol. Cell. Biol.* **3**, 2156–2165.
- Stevens, R. C. (2004). *Nature Struct. Mol. Biol.* **11**, 293–295.
- Törnroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. & Kjellborn, P. (2005). *Nature (London)*, **439**, 688–94.
- Turnbull, A. P., Kummel, D., Prinz, B., Holz, C., Schultchen, J., Lang, C., Niesen, F. H., Hofmann, K.-P., Delbruck, H., Behlke, J., Muller, C., Jarosch, E., Sommer, T. & Heinemann, U. (2005). *EMBO J.* **24**, 875–884.
- Vialard, J., Lalumiere, M., Vernet, T., Briedis, D., Alkhatib, G., Henning, D., Levin, D. & Richardson, C. (1990). *J. Virol.* **64**, 37–50.
- Wu, H., Kuong, P. D. & Hendrickson, W. A. (1997). *Nature (London)*, **387**, 527–530.
- Xu, X. & Jones, I. M. (2004). *Virus Genes*, **29**, 191–197.
- Yao, Y., Ren, J. Y., Heinen, P., Zambon, M. & Jones, I. M. (2004). *J. Infect. Dis.* **190**, 91–98.
- Zhao, Y., Chapman, D. A. & Jones, I. M. (2003). *Nucleic Acids Res.* **31**, E6.