A time- and cost-efficient system for high-level protein production in mammalian cells

A. Radu Aricescu, Weixian Lu and E. Yvonne Jones*

Cancer Research UK Receptor Structure Research Group, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, England

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

Most proteins for structural biology studies are produced by high-level expression in *Escherichia coli*. However, prokaryotic based expression systems fail to generate correctly folded functional forms of many proteins and hence a variety of eukaryotic based expression systems have been developed. Of these, yeast and baculovirus-infected insect cells currently represent the expression systems of choice for structural biologists. Here, protocols for a simple, fast and affordable method for transient protein expression in mammalian cells are reported. The results demonstrate that it combines several features necessary for the production of suitable samples for structural biology, in particular protein crystallography, namely high protein yield, straightforward purification, selenomethionine incorporation and control of N-linked glycosylation. The system is suitable for use in conventional laboratories or can be implemented in a medium- or high-throughput pipeline.

1. Introduction

Intuitively, a protein will be best expressed in its native cell type under physiological conditions, where a multitude of molecular systems work together for efficient production and quality control at various stages, including synthesis and folding, post-translational modifications and subcellular targeting. However, only a small number of proteins occur naturally in amounts that permit convenient purification of the relatively large quantities required for structural studies. Therefore, various heterologous overexpression systems have been developed to produce recombinant proteins: bacterial (most commonly *Escherichia coli*), yeast (*Pichia pastoris* and *S. cerevisiae*), baculovirus-infected insect cells, mammalian cells and, more recently, cell-free systems (reviewed in Yokoyama, 2003).

Analysis of the Protein Data Bank shows that >90% of proteins used for structural studies have been expressed in *E. coli*, where the procedures are honed to provide a reliable, fast and cheap system. This remains true today for high-throughput (HTP) studies, despite the increasing emphasis on the structural biology of proteins which are not necessarily amenable to prokaryotic expression; for example, the human and viral protein targets of the Structural Proteomics In Europe (SPINE) consortium (see Alzari *et al.*, 2006; Aricescu, Assenberg *et al.*, 2006; Banci *et al.*, 2006; Fogg *et al.*, 2006). However, satisfactory expression of eukaryotic proteins in a stable and soluble form is a major bottleneck and routine, cost-effective and quick systems for expression in eukaryotic cells would be very valuable. The technology developments undertaken as part of various structural genomics programmes have so far failed to achieve this in practice (Stevens,

research papers

2004; Aricescu, Assenberg *et al.*, 2006). Yeast systems, although cheap, scaleable and suitable for HTP processing, have in general proved of value mainly for small (<500 residues) cytoplasmic proteins. Furthermore, there is little evidence that the overall success rate is significantly higher than for *E. coli*. However, the obvious alternatives, insect and mammalian systems, are usually seen as slow, expensive and technically more challenging.

A mammalian expression system would be the ideal choice to support structural studies focused on the characterization of human cell-surface receptors and secreted proteins, since efficient post-translational processing in the secretory pathway is crucial for the correct folding of such proteins. Secreted mammalian proteins have been successfully produced in a large-scale format in Chinese hamster ovary (CHO) cells (using a system initially developed by Cockett *et al.*, 1990) and human embryonic kidney (HEK) 293EBNA cells (Meissner *et al.*, 2001; Durocher *et al.*, 2002; Geisse & Henke, 2005). Each method has its own advantages: protocols for efficient selenomethionine (SeMet) incorporation and reduced N-glycosylation in the case of CHO cells (Jones *et al.*, 1992; Butters *et al.*, 1999; Davis *et al.*, 2001); transient transfection protocols for 293EBNA cells grown in suspension cultures. Each system also has disadvantages: for CHO cells it is the long time required to obtain stable clones that produce sufficient amounts of protein, whereas 293EBNA cells require specialized (and expensive) media for a successful transfection (Geisse & Henke, 2005). Mancia *et al.* (2004) have recently described an ingenious method for selecting high-level protein producing clones in stably transfected HEK293T cells. However, this method remains rather laborious and time-consuming.

We present here a simple, cost-effective and fast mammalian cell-based expression system that will not only support small-scale screening of a large number of constructs in parallel, but will also allow rapid and reliable scale-up for milligram-quantity production for X-ray crystallography experiments. The system also allows the protein to be labelled with SeMet, gives control over the level of N-linked glycosylation and provides an effective route to the expression of macromolecular complexes. This should provide the basis for development of HTP expression screening platforms using mammalian cells.

2. Materials and methods

2.1. Vector design

The main criteria that were taken into account in choosing the plasmid backbone for our expression vectors were as follows: the plasmid should have a very high copy number in *E. coli* (for efficient DNA production), its promoter should be the strongest available for mammalian cells and, finally, it should be small enough to allow the efficient cloning of variable-length constructs. Following these guidelines, we chose the pCAβ-EGFP backbone (kindly provided by Jonathan Gilthorpe, King's College London; this is a derivative of

the pCAGGS plasmid developed in J.-I. Miyazaki's laboratory, Osaka University Medical School; Niwa *et al.*, 1991; Fukuchi *et al.*, 1994). The EGFP cDNA and multiple cloning site of pCAβ-EGFP were removed and replaced with a linker containing the following unique restriction sites: *Bsm*BI, *Eco*RI, *Kpn*I, *Sac*I, *Not*I, *Sph*I and *Xho*I. The resulting vector was termed pLEXm (to differentiate it from the pLEX vector commercialized by Invitrogen, which is only suitable for bacterial expression). pLEXm contains the pBR322 origin of replication, ampicillin resistance, a cytomegalovirus enhancer and chick β-actin promoter (together, these elements were found to be the strongest out of nine promoters tested by Fukuchi *et al.*, 1994, in a variety of cell lines) as well as the rabbit β-globin intron and poly-A signal.

Building on the pLEXm backbone, we developed a series of vectors containing several tags for purification and detection [His_8, Lys-His_6, influenza hemagglutinin-HA-epitope, the human IgGγ1 Fc region and the biotin ligase (BirA) recognition site], optimized Kozak and secretion signal sequences and variations of the multiple cloning site. These vectors include pHLsec, in which inserts can be cloned between a secretion signal sequence and a Lys-His_6 tag, designed to suitable for constructs lacking N-terminal domains, pHL-FcHis, which allows expression of a C-terminal Fc fusion protein tag, and pHL-Avitag3, which is suitable for surface plasmon resonance experiments since the constructs are fused with a 15-residue C-terminal biotinylation tag. Maps for all these vectors are provided as supplementary information[1] and the vectors are available on request.

2.2. DNA purification

We purify DNA using the Endotoxin-Free Plasmid Mega Kit (Qiagen). High-quality DNA is essential for successful transfection and we found that only samples with an OD_{260}/OD_{280} ratio of 1.8 or higher are suitable.

The pLEXm and derivative plasmids are all very high copy number, therefore one can expect between 2 and 4 mg of pure DNA from a 500 ml overnight bacterial culture (in our experience, cultures grown for more than 12–14 h give poor yields).

The DNA samples must be sterile; therefore, care should be taken to wash the DNA precipitates properly with 70% ethanol before dissolving them in sterile 10 m*M* Tris pH 8.

2.3. Cell lines

We chose HEK293T cells as our main cell line owing to ease of handling, robust growth rate, excellent transfectability, high capacity for recombinant protein expression and low-cost media requirements. They are available from all major cell banks and support repeated passages (twice a week for at least six months) without losing the above properties. These cells also tolerate media low in serum or supplemented with SeMet very well (see below).

---

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5088). Details for accessing this material are given at the back of the journal.

N-Acetylglucosaminyltransferase I-negative 293S GnTI⁻ cells (Reeves *et al.*, 2002), unable to synthesize complex N-glycans, were a kind gift from Philip Reeves and H. Gobind Khorana (Massachusetts Institute of Technology).

Both cell lines can be maintained in Dulbecco's Modified Eagle's Medium (DMEM high glucose, Sigma) supplemented with L-glutamine, non-essential amino-acids (Gibco) and 10% foetal calf serum (FCS, Sigma). Cells are normally grown in standard flasks (Nunc) in a humidified 310 K incubator with 5% CO_2. Large-scale cultures for protein production are performed in expanded-surface polystyrene roller bottles (2125 cm^2, Greiner Bio-One). These roller bottles do not have a gas-permeable cap; therefore, it is important to flow CO_2 gas into each bottle for 20–30 s before tightening the cap.

2.4. Transfection reagent

Polyethylenimine (PEI) is a very affordable and highly efficient transfection reagent and there is now a significant

Day 1: DNA cloning

↓

Day 3: Plasmid QC (restriction, sequencing)

↓

Day 4: Small-scale transfection

↓

Day 7: Expression test via dot or Western blot

↓

Day 8: Scale-up plasmid DNA of positive clones

↓

Day 9: Large-scale transfection in roller bottles

↓

Day 13: Conditioned medium is collected; IMAC; overnight deglycosylation if necessary

↓

Day 14: Gel filtration and crystallization trials

Figure 1
Flow chart of the transient mammalian expression system. The whole procedure, from target cloning to crystallization trials, can be completed in two weeks.

number of publications reporting its use (reviewed by Kichler, 2004; Demeneix & Behr, 2005). A large selection of PEI forms are available (Aldrich) varying in molecular weight and branching. The PEI we use is '25 kDa branched', which was found to be most effective in transfecting 293EBNA cells by Durocher *et al.* (2002). Stock solutions are made in water, first at 100 mg ml^{-1} (PEI is an extremely viscous liquid, which cannot be pipetted). Once the solution is homogeneous, it is further diluted to 1 mg ml^{-1}, the pH adjusted to 7 with HCl, filter sterilized and aliquoted. Aliquots are stored frozen and working solutions can be kept at 277 K for long periods of time (months) without any loss in transfection efficiency.

2.5. Transfection protocol

Optimal transfection is achieved when adherent cells reach about 90% confluency. For a standard 15 cm tissue-culture dish (approximate surface area 176 cm^2), 50 µg of plasmid DNA are required (quantities need to be scaled accordingly for the different surface areas of plates/flasks/bottles). The DNA is added to 5 ml serum-free medium and mixing is followed by the addition of 75 µl PEI stock (1 mg ml^{-1}) and brief vortexing. The solution is incubated for 10 min at room temperature to allow DNA–PEI complex formation. In pilot experiments, we found that the optimal DNA:PEI ratio is between 1:1.5 and 1:2; however, we prefer 1:1.5 owing to lower
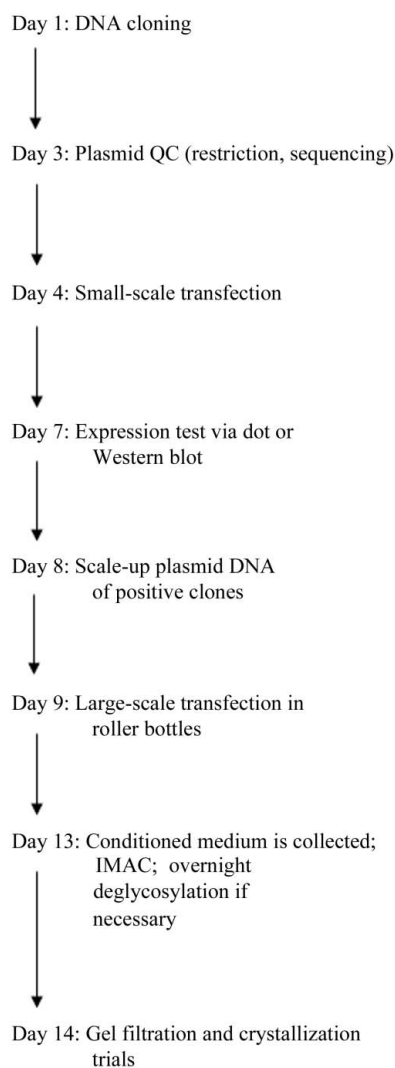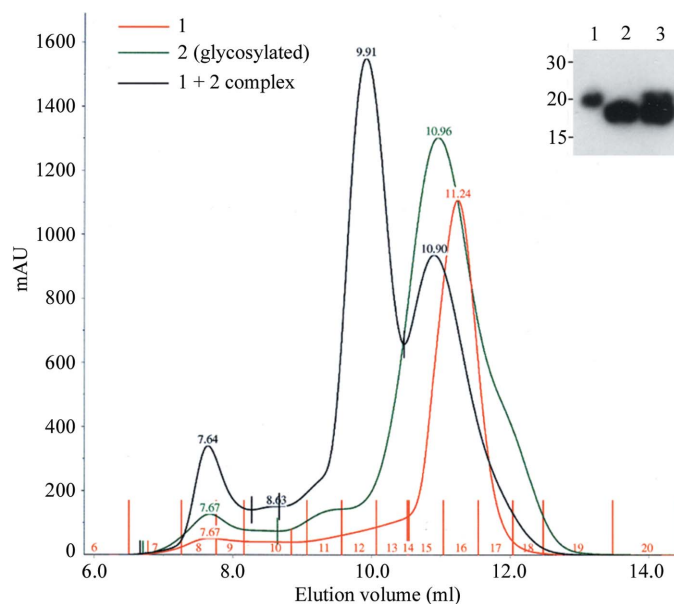

Figure 2
Efficient production of a receptor–ligand complex by co-transfection in HEK293T cells. Cells were transfected with either the receptor construct (sample 1) or the ligand (sample 2, N-glycosylated) or a 1:1 mix of the two plasmids (1 + 2 complex and lane 3 in the inset). Protein levels appear to be improved by co-expression, considering that only half of the normal amount of vector DNA for each partner was used. Size-exclusion chromatography demonstrates complex formation and reveals that a significant amount of ligand (expressed normally at higher levels than the receptor) remains unbound. In the chromatogram, sample 2 appears larger than sample 1 owing to N-glycosylation. For the Western blot analysis (inset) the samples first were deglycosylated by treatment with PNGase; they therefore migrate at the predicted molecular weights.

toxicity. During complex formation, media from the plates to be transfected should be changed, lowering the serum concentration to 2% (20 ml of fresh medium for the 15 cm dish taken as an example here). Finally, the DNA–PEI complex is added to the dish, which is then briefly rotated to allow mixing and the cells are placed in the incubator. 3–4 d later, conditioned medium is ready for collection and protein purification.

2.6. Western blotting

For the rapid screening of new constructs, transfections are performed in 24-well or 6-well plates using MiniPrep-purified DNA (QIAprep Spin Kit, Qiagen). Plasmid quantities required per well are 1 or 5 µg, respectively. 3 d post-transfection, small aliquots of conditioned media (5–10 µl) are analysed by Western blotting using the PentaHis monoclonal primary antibody (1:1000 dilution, Qiagen) and goat anti-mouse IgG peroxidase-conjugated secondary antibody (1:2000, Sigma). The signal can be visualized by chemi-luminescence using the ECL kit (GE Healthcare).

2.7. Purification methods

Recombinant proteins are readily purified from conditioned media, with protocols that vary depending on the affinity tags used. We normally purify His-tagged constructs from large-scale cultures using the following IMAC batch procedure. Conditioned media are filtered through a 0.2 µm membrane (Express filter, Millipore), diluted threefold with PBS and the final pH is adjusted to 8 by adding Tris buffer to a final concentration of 10 mM. IMAC purification is performed using nickel-coated chelating Sepharose (GE Healthcare) or cobalt-coated Talon beads (Clontech). A final round of gel filtration is usually sufficient to yield proteins of >90% purity suitable for crystallization trials.

2.8. Selenomethionine labelling

Cells are grown using the standard conditions described in §2.3 prior to transfection. During transfection (see §2.5), the following modifications to the protocol are implemented: the cells are washed once with phosphate-buffered saline (PBS) before adding fresh medium containing 2% serum in order to remove most of the residual methionine; the fresh medium, as well as the serum-free medium used for DNA–PEI complex formation, is methionine-free DMEM (MP Biomedicals) supplemented with L-glutamine, non-essential amino acids and 30 mg l^{-1} SeMet (Acros Organics). The serum used must be dialysed (available from Sigma) to remove any traces of free methionine. 3–4 d post-transfection, the conditioned medium is ready for collection. A significant amount of cell death is normal owing to the toxicity of SeMet, which leads to reduced levels of protein expression.

3. Results and discussion

3.1. Timeframe considerations

There is a considerable time pressure in modern structural biology, most explicitly in so-called HTP pipelines. The methods described here are far faster than the 2–3 months required for systems using stable transfection of mammalian cells (Cockett et al., 1990; Mancia et al., 2004).

Typically, 3 d are required for the initial small-scale expression screening test. Once well expressing constructs have been identified, the DNA scale-up takes 1 d and the process from large-scale transfection to protein purification and set-up of crystallization trials can be accomplished in less than a week (Fig. 1). These timescales mean that compared with other eukaryotic expression systems the mammalian transient transfection method is significantly faster than the baculovirus-infected insect-cell system or generation of yeast stable integrants (timescale considerations discussed in Mancia et al., 2004; Aricescu, Assenberg et al., 2006).

3.2. Transfection: strengths and weaknesses

A disadvantage of the transient method is the fact that new cells have to be transfected for every batch of protein required (although it is possible to collect two batches of protein from each transfection by refreshing the medium after 3–4 d; in our experience, cells will keep secreting proteins for 7–10 d; data not shown). On the other hand, a significant advantage of the system is that co-transfection experiments can be easily performed, providing an excellent vehicle for the high-level expression of protein complexes (Fig. 2).

3.3. Throughput and protein yield

Expression screening tests can be easily performed in a medium- to high-throughput format (for example, 24 constructs at a time; shown in Fig. 3). The Western blot procedure can be calibrated using known amounts of protein markers and reliable predictions regarding the scale-up yields can be made from these small-scale screens (Fig. 4a). The protein yields obtained, based on the set of 24 constructs shown in Fig. 3, range between 1 and 40 mg of final pure protein (ready for crystallization) per litre of culture (which requires four roller bottles). This is sufficient for screening a large number of crystallization conditions using nanolitre robotic systems (Walter et al., 2003, 2005). We have not observed a correlation between the expression level and the protein size (constructs tested cover a size range of 20–120 kDa), number of disulfide bonds or the amount of N-linked glycosylation. We have successfully expressed secreted proteins containing domains with various folds (Ig-like, FNIII-like, jelly roll, seven-blade propeller, leucine-rich repeats). Recent examples include the MAM-Ig N-terminal domains of the receptor protein tyrosine phosphatase mu (Aricescu, Hon et al., 2006) and semaphorin 4D (originally expressed in stably transfected CHO-Lec3.2.8.1 cells; Love et al., 2003). However, most of the successfully expressed constructs represent fragments of or full-length extracellular

regions of cell-surface receptors. Membrane and intracellular proteins can also be overexpressed in HEK293T cells (Fig. 4b and 4c) although, based on a very limited set of targets tested so far, the yields appear to be lower (fivefold to tenfold) compared with the baculovirus expression system. Never-theless, small-scale transient transfections represent a very convenient way of testing new constructs before initiating baculovirus production. Further improvements will be required (more powerful vectors, development of efficient protocols for maintenance and transfection of cells in suspension) for the routine production of large amounts of membrane and intracellular proteins in mammalian cells.

Another factor with a huge influence on the recombinant protein yields is (in any system) a careful design of the constructs to be expressed. Failure to express can often be linked to failures in the use of various bioinformatics tools (for example, errors in defining domain boundaries). Also, it may be that other cell lines are more suitable owing to particular processing requirements: appropriate co-translational glycosyl-ation can be essential for correct folding and expression of the construct (Wormald & Dwek, 1999) and co-transfection with an interaction partner can be essential for efficient secretion or increased stability (Coe et al., 2001).

Recently, we found that HEK293T cells tolerate a sudden serum depletion post-transfection surprisingly well (Fig. 5). Even the complete removal of foetal calf serum (FCS) from the medium added post-transfection does not appear to significantly reduce the yields of secreted proteins, thus removing the only major source of protein contamination and opening up the possibility of setting up crystal-lization screens following a highly simplified purification and concentra-tion protocol (see Fig. 5a; note that residual bovine serum albumin and other proteins, possibly secreted by cells or resulting from dead cells, are still present but in very small amounts).

Figure 3
Expression test of a set of 24 extracellular constructs produced by small-scale transfection. Western blot analysis of 3 d conditioned media revealed that 18 out of the 24 constructs tested (which vary in size, glycosylation state and domain composition) can be detected. The detection threshold corresponds to approximately 1 mg l^{-1} protein of interest in the conditioned medium. Molecular-weight markers values are in kDa.
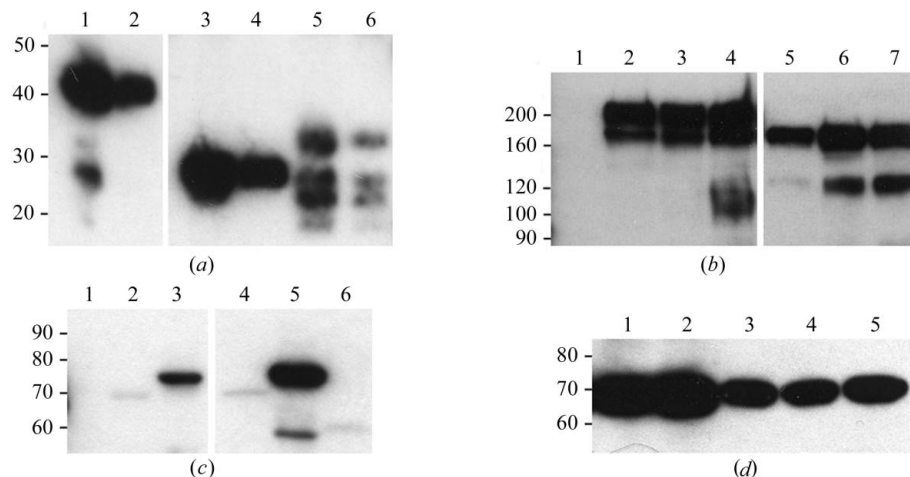
Figure 4
Calibration tests for protein expression in eukaryotic cells. All panels are Western blots probed with an anti-His-tag monoclonal antibody (see §2.6). (a) Small versus large-scale expression: three constructs were transfected either in six-well dishes (lanes 1, 3 and 5) or in roller bottles (lanes 2, 4 and 6) and samples of conditioned media collected 3 d later. A higher concentration of protein in the conditioned media was obtained in the small-scale experiments, but the ratios between small-scale and large-scale yields appear constant, thus allowing a good estimate of the scale-up required. (b) Expression of a large (170 kDa) membrane protein in HEK293T (lanes 2–4, transient transfection) and Sf9 (lanes 5–7, baculovirus infection) cells. Lane 1 is negative control (non-transfected HEK293T cells). Samples containing similar amounts of cells were collected 24 h (lanes 2 and 5), 48 h (lanes 3 and 6) and 72 h (lanes 4 and 7) post-transfection/infection. The 200 kDa band observed in the HEK cells samples arises from more complex glycosylation. A breakdown product of ~120 kDa becomes apparent as cultures grow older. (c) Parallel expression test of two intracellular proteins in HEK293T (lanes 2 and 3) and Sf9 (lanes 4 and 5) cells. Lanes 1 and 6 are negative controls (non-transfected/non-infected HEK293T/Sf9 cells, respectively). One construct (lanes 2 and 4) is poorly expressed in both systems, while the second construct (lanes 3 and 5) appears to be better expressed in insect cells. (d) SeMet addition to the tissue-culture media, post-transfection, causes a reduction in protein yield. Lane 1, complete medium; lane 2, Met-free medium supplemented with 30 mg l^{-1} L-methionine; lanes 3, 4 and 5, Met-free medium supplemented with 30 mg l^{-1} SeMet (three different batches).

3.4. Cost and technology considerations

Traditionally, mammalian cell expression systems have been seen as expensive owing to the costs associated with transfection reagents, specialized media, consumables and the expert workforce required. However, the cost of PEI reagent per litre of transfected cells is very close to zero. HEK293T cells are very easy to maintain and they
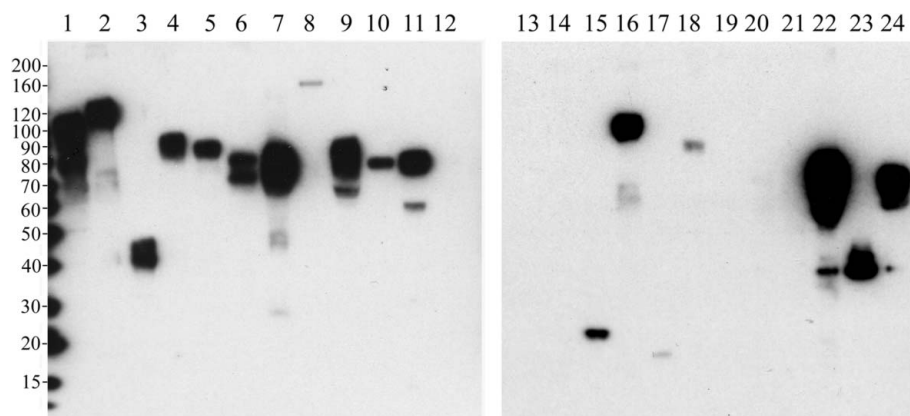
research papers

grow in standard DMEM, one of the cheapest eukaryotic cell media available. The costs associated with plasticware have already been reduced by the use of expanded-surface roller bottles and the current competition between suppliers is likely to drive prices down further. Proper training in the use of sterile techniques and good technical support are indeed crucial, but we would argue that these are skills worth investing in. Several steps in our protocols can be performed by robots (plasmid DNA purification, small-scale transfection of cells for expression testing) and it is possible that in the near future more steps will be optimized for automation (N. Berrow, personal communication).

3.5. Selenomethionine labelling

In vivo SeMet incorporation into recombinant proteins can be performed in virtually all expression systems and together with the multi-wavelength anomalous dispersion (MAD) method it has revolutionized crystallographic phasing (Hendrickson *et al.*, 1990; Bellizzi *et al.*, 1999; Laurila *et al.*, 2005; for a discussion, see Aricescu, Assenberg *et al.*, 2006). The first report of successful labelling and structure determination of an SeMet recombinant protein expressed in mammalian cells (CHO; Lustbader *et al.*, 1995) and subsequent simplification of the method (May *et al.*, 1997; Davis *et al.*, 2001) encouraged us to try applying it in our transient transfection setup using HEK293T cells.

We observed a drop in expression levels of approximately 50% (Fig. 4*d*) when media are supplemented with SeMet, mainly owing to cell death caused by SeMet toxicity. Nevertheless, in favourable cases, once crystallization conditions for the native protein have been established, obtaining crystals of the SeMet-labelled construct may require little optimization and therefore modest amounts of protein. Here, we present two examples (Fig. 6) where structures were solved using SeMet proteins produced in HEK293T cells. The incorporation was checked directly at the synchrotron (beamline BM14, ESRF, Grenoble) by recording Se *K*-edge fluorescence scans of the crystals. Mass-spectrometry analysis of the protein samples revealed that approximately 60% of the purified protein was labelled, in good agreement with the refined occupancies derived from the program *SOLVE* (Terwilliger, 2003) for the various Se atoms: 0.42, 0.47 and 0.58 for example 1 (three Se atoms per molecule), and 0.87 and 0.68 for example 2 (two Se atoms per molecule). The signal obtained (for roughly one Met residue per 100 amino acids in both cases) was sufficient to solve the structure in both cases (Aricescu, Hon *et al.*, 2006, and manuscript in preparation).

3.6. Control of N-linked glycosylation

Mammalian cells (including HEK293T) synthesize complex N-glycans and this is likely to affect the ability of the recombinant proteins to form well diffracting crystals (Davis *et al.*, 1993). Apart from the increased solubility of the glycosylated proteins and the inherent flexibility of large and branched sugar chains, it is likely that certain sites will only be occupied in a subpopulation of the molecules, in part owing to overexpression overwhelming the cells' processing capacity, and this can introduce a source of heterogeneity in the protein samples (see Fig. 2, where most proteins are N-glycosylated and appear as multiple bands in the gel).

However, the presence of sugars may be strictly required for correct folding of some proteins. For example, we have observed a case in which a protein normally secreted at very high levels was retained in the ER and degraded when cells were treated with tunicamycin, an inhibitor of N-glycosylation (data not shown). There are two obvious lines of attack on this problem, the use of mutant cells and the use of drugs to interfere with glycosylation; both of these can be very effective (Davis *et al.*, 1993; Butters *et al.*, 1999). In the context of the present protocols a reasonable compromise is the use of the 293S GnTI⁻ cell line developed in Khorana's laboratory (Reeves *et al.*,
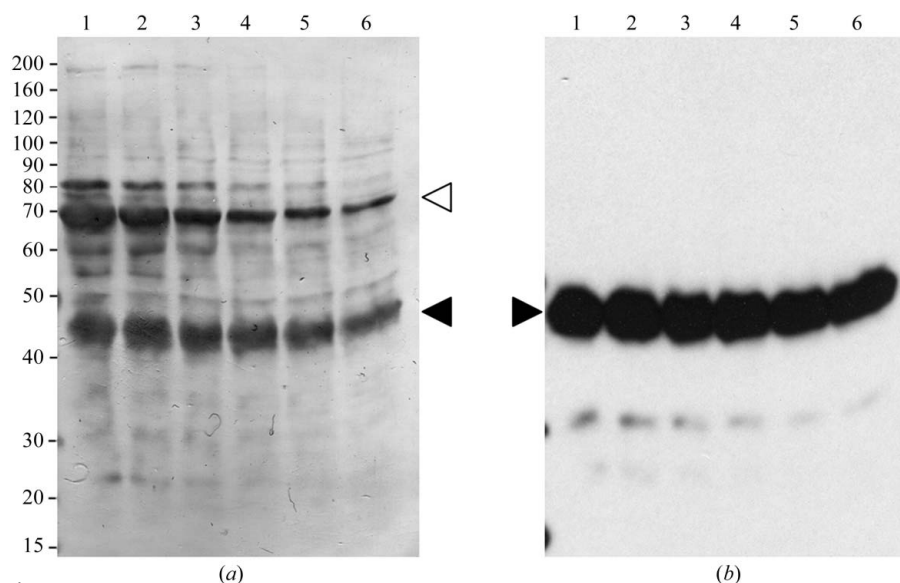
Figure 5
Influence of serum depletion on the recombinant protein yield. (*a*) Immediately following PEI transfection, cells were grown for 3 d in media supplemented with 2% (lane 1), 1% (lane 2), 0.5% (lane 3), 0.2% (lane 4), 0.1% (lane 5) and 0% (lane 6) serum. Equal amounts of conditioned media were separated by SDS–PAGE and transferred on a nitrocellulose membrane. The membrane was stained with Coomassie Brilliant Blue G-250 after being probed with anti-His antibodies. The overexpressed protein of interest is indicated by a black arrowhead. Bovine serum albumin, the major contaminant from the serum, is indicated by an open arrowhead. Reducing serum levels clearly removes the majority of the contaminant proteins in the medium. However, residual BSA is still present as well as small amounts of some other proteins, probably secreted by the cells. (*b*) Western blot analysis of the conditioned media confirms that protein yields remain unaffected despite the sudden serum depletion.

2002) that produces very homogeneous and lower molecular-weight sugar chains (Man$_5$GlcNAc$_2$). A detailed characterization of proteins produced in this cell line will be reported separately (Chang *et al.*, manuscript in preparation) and
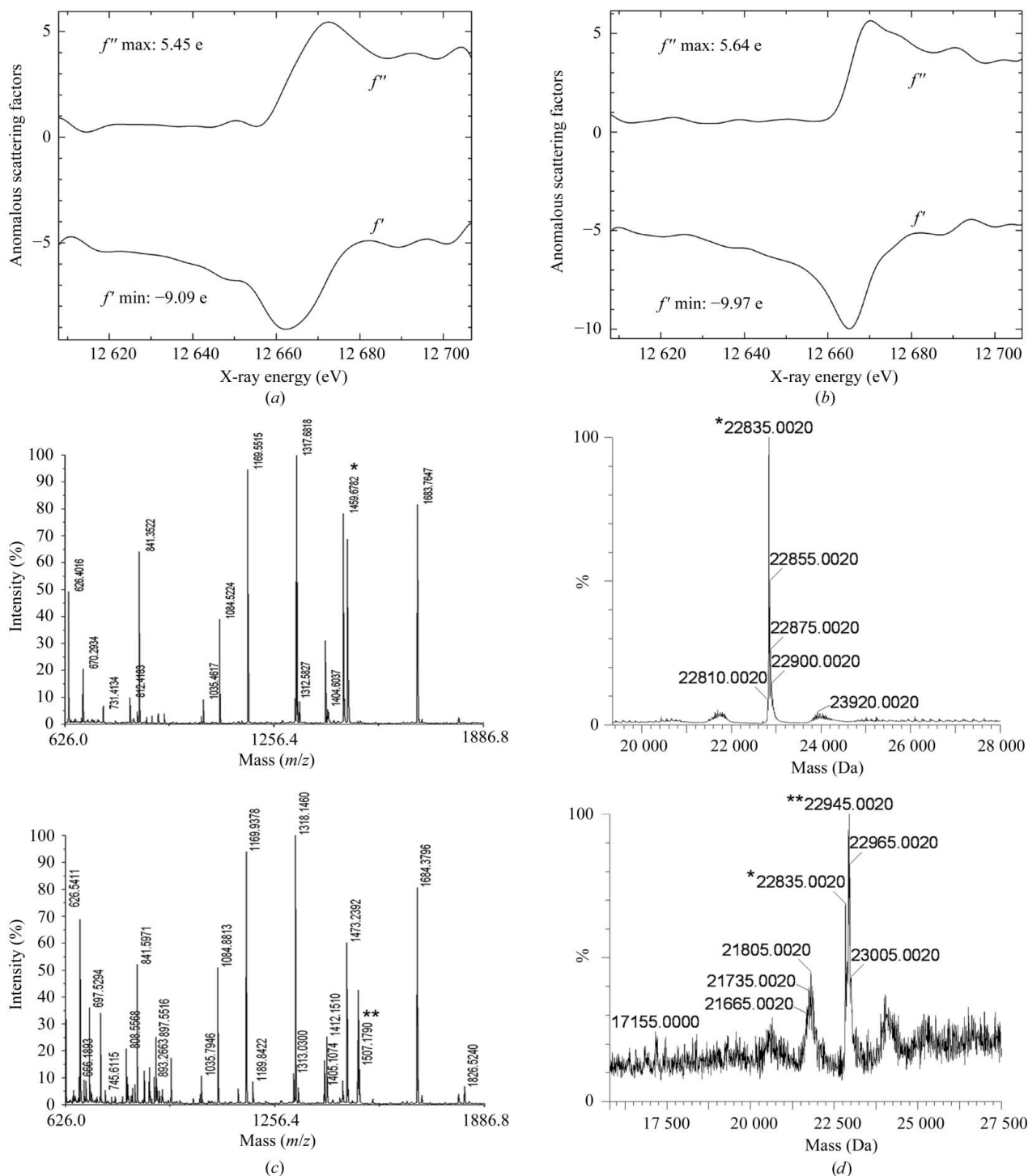

Figure 6
Analysis of SeMet incorporation in two proteins expressed in mammalian cells. (*a*) and (*c*) correspond to one sample and (*b*) and (*d*) to the second. (*a*) and (*b*) The two constructs were crystallized and Se *K*-edge fluorescence scans performed at ESRF beamline BM14 demonstrate Se incorporation. (*c*) Mass-spectrometric analysis of the peptides obtained by trypsin digestion of the first sample in its native and SeMet-labelled states. The observed pattern is essentially identical apart from one peptide (single star) that is shifted in the SeMet-labelled sample (double star) by a mass corresponding to one sulfur-to-selenium substitution. Two other peptides containing Met residues in this sample could not be detected. However, labelling of these Met residues was subsequently confirmed by crystallographic data processing. (*d*) Mass-spectrometric analysis of the second protein sample, undigested, containing two Met residues, reveals a peak shift corresponding to teh incorporation of two Se atoms (see double-star *versus* single-star labels). The data indicate that approximately 60% of the protein sample was labelled.

further strategies to reduce N-linked glycosylation are discussed in Aricescu, Assenberg *et al.* (2006).

4. Conclusions

The method described here allows high-level expression of secreted proteins in mammalian cells in a quantity and of a quality suitable for X-ray crystallography. It is fast, cheap, reliable, easy scalable and can be used for high-throughput experiments. It provides an easy route to the co-expression of components of protein complexes and we have adapted existing methods to allow SeMet incorporation. The use of an N-glycosylation-deficient cell line allows production of more homogeneous samples of glycoproteins. Further developments are required in order to allow efficient expression of cytoplasmic and membrane proteins as well as to further manipulate the N- and O-linked glycosylation.

References

Alzari, P. M. *et al.* (2006). *Acta Cryst.* D**62**, 1103–1113.
Aricescu, A. R., Assenberg, R. *et al.* (2006). *Acta Cryst.* D**62**, 1114–1124.
Aricescu, A. R., Hon, W. C., Siebold, C., Lu, W., van der Merwe, P. A. & Jones, E. Y. (2006). *EMBO J.* **25**, 701–712.
Banci, L. *et al.* (2006). *Acta Cryst.* D**62**, 1208–1217.
Bellizzi, J. J., Widom, J., Kemp, C. W. & Clardy, J. (1999). *Structure*, **7**, R263–R267.
Butters, T. D., Sparks, L. M., Harlos, K., Ikemizu, S., Stuart, D. I., Jones, E. Y. & Davis, S. J. (1999). *Protein Sci.* **8**, 1696–1701.
Cockett, M. I., Bebbington, C. R. & Yarranton, G. T. (1990). *Biotechnology*, **8**, 662–667.
Coe, A. P., Askari, J. A., Kline, A. D., Robinson, M. K., Kirby, H., Stephens, P. E. & Humphries, M. J. (2001). *J. Biol. Chem.* **276**, 35854–35866.
Davis, S. J., Ikemizu, S., Collins, A. V., Fennelly, J. A., Harlos, K., Jones, E. Y. & Stuart, D. I. (2001). *Acta Cryst.* D**57**, 605–608.
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.
Demeneix, B. & Behr, J. P. (2005). *Adv. Genet.* **53**, 217–230.
Durocher, Y., Perret, S. & Kamen, A. (2002). *Nucleic Acids Res.* **30**, E9.
Fogg, M. *et al.* (2006). *Acta Cryst.* D**62**, 1196–1207.
Fukuchi, K., Hearn, M. G., Deeb, S. S., Smith, A. C., Dang, N., Miyazaki, J., Bothwell, M. & Martin, G. M. (1994). *In Vitro Cell Dev. Biol. Anim. A*, **30**, 300–305.
Geisse, S. & Henke, M. (2005). *J. Struct. Funct. Genomics*, **6**, 165–170.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). *Nature (London)*, **360**, 232–239.
Kichler, A. (2004). *J. Gene Med.* **6**, Suppl. 1, S3–S10.
Laurila, M. R., 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.
Lustbader, J. W., Wu, H., Birken, S., Pollak, S., Gawinowicz Kolks, M. A., Pound, A. M., Austen, D., Hendrickson, W. A. & Canfield, R. E. (1995). *Endocrinology*, **136**, 640–650.
Mancia, F., Patel, S. D., Rajala, M. W., Scherer, P. E., Nemes, A., Schieren, I., Hendrickson, W. A. & Shapiro, L. (2004). *Structure*, **12**, 1355–1360.
May, A. P., Robinson, R. C., Aplin, R. T., Bradfield, P., Crocker, P. R. & Jones, E. Y. (1997). *Protein Sci.* **6**, 717–721.
Meissner, P., Pick, H., Kulangara, A., Chatellard, P., Friedrich, K. & Wurm, F. M. (2001). *Biotechnol. Bioeng.* **75**, 197–203.
Niwa, H., Yamamura, K. & Miyazaki, J. (1991). *Gene*, **108**, 193–199.
Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 13419–13424.
Stevens, R. C. (2004). *Nature Struct. Mol. Biol.* **11**, 293–295.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Walter, T. S., Diprose, J. M., Brown, J., Pickford, M. G., Owens, R. J., Stuart, D. I. & Harlos, K. (2003). *J. Appl. Cryst.* **36**, 308–314.
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.* D**61**, 651–657.
Wormald, M. R. & Dwek, R. A. (1999). *Structure*, **7**, R155–R160.
Yokoyama, S. (2003). *Curr. Opin. Chem. Biol.* **7**, 39–43.

Aricescu, A.R., Lu, W. and Jones, E.Y. "A time and cost efficient system for high level protein production in mammalian cells"

Supplementary information: vector maps
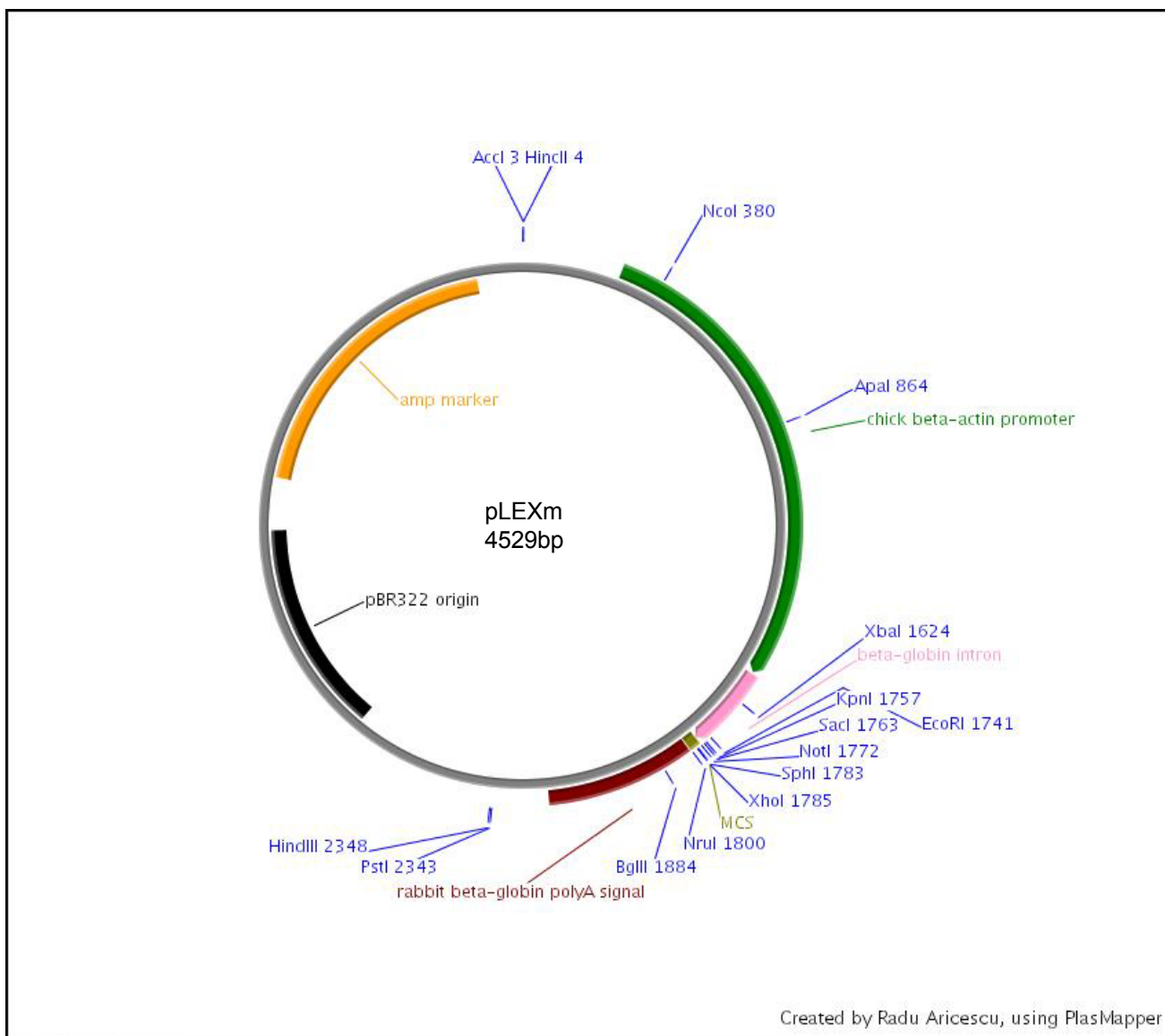
MCS:

For **secreted constructs**, a vector termed **pHLsec** is available. It is based on the pLEXm backbone, the MCS however was reduced when several new features were introduced:

-a **Kozak sequence**;
-a **secretion signal sequence**;
-a **C-terminal K_His6 tag**.

Constructs can be cloned into this vector, preserving and making use of the features listed above, by using the **AgeI** (compatible with XmaI, BspEI, NgoM IV) and **KpnI** (isoschizomer Acc65I is compatible with BsiWI and BsrGI) sites.

```
EcoRI HindIIIKozak M  G  I  L  P  S  P  G  M  P  A  L  L  S
GAATTCAAGCTTGCCACCATGGGGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCC

 L  V  S  L  L  S  V  L  L  M  G  C  V  A  E  T  G
CTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGTAGCTGAAACCGGT...insert...

 G  T  K  H  H  H  H  H  H  *  *   XhoI
GGTACCAAGCACCACCATCACCATCACTAATGATCACTCGAG
```

Fc-tagged constructs, allowing purification of fusion proteins on Protein A affinity columns, can be expressed using the **pHL-FcHis** vector.
The vector is based on the pLEXm backbone and it contains a **3C protease cleavage site** followed by the **human IgGγ1 hinge and Fc** regions and finally a **KHis6** tag, all cloned between the **KpnI** site and the **XhoI** site.

Constructs can be cloned into this vector, preserving and making use of the features listed above, by using **EcoRI** (compatible with MfeI) and **KpnI** (isoschizomer Acc65I is compatible with BsiWI and BsrGI) sites or by ligation-independent cloning.

```
EcoRI
GAATTC .....................insert..........................

KpnI
 G   T   L   E   V   L   F   Q   G   P   K   S   C   D   K   T   H   T   C   P
GGTACCCTGGAGGTGCTGTTCCAGGGCCCCAAATCTTGTGACAAAACTCACACATGCCCA
 P   C   P   A   P   E   L   L   G   G   P   S   V   F   L   F   P   P   K   P
CCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCC
 K   D   T   L   M   I   S   R   T   P   E   V   T   C   V   V   V   D   V   S
AAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGC
 H   E   D   P   E   V   K   F   N   W   Y   V   D   G   V   E   V   H   N   A
CACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCC
 K   T   K   P   R   E   E   Q   Y   N   S   T   Y   R   V   V   S   V   L   T
AAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACC
 V   L   H   Q   D   W   L   N   G   K   E   Y   K   C   K   V   S   N   K   A
GTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCC
 L   P   A   P   I   E   K   T   I   S   K   A   K   G   Q   P   R   E   P   Q
CTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAG
 V   Y   T   L   P   P   S   R   D   E   L   T   K   N   Q   V   S   L   T   C
GTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCTGACCTGC
 L   V   K   G   F   Y   P   S   D   I   A   V   E   W   E   S   N   G   Q   P
CTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCG
 E   N   N   Y   K   A   T   P   P   V   L   D   S   D   G   S   F   F   L   Y
GAGAACAACTACAAGGCCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTAC
 S   K   L   T   V   D   K   S   R   W   Q   Q   G   N   V   F   S   C   S   V
AGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTG
 M   H   E   A   L   H   N   H   Y   T   Q   K   S   L   S   L   S   P   G   K
ATGCATGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAA
 H   H   H   H   H   H   *   *      XhoI
CACCACCATCACCATCACTAATGATCACTCGAG
```

Supplementary information: Page 3

Biotinylated constructs (for surface plasmon resonance binding experiments for example, see Aricescu *et al.*, 2006) can be expressed using the vector termed **pHL-Avitag3**.

The vector is based on the pLEXm backbone and it contains a **GGS linker region** followed by the **biotin ligase [BirA] recognition site** and finally a **KHis6** tag, all cloned between the **KpnI** site and the **XhoI** site.

Constructs can be cloned into this vector, preserving and making use of the features listed above, by using **EcoRI** (compatible with MfeI) and **KpnI** (isoschizomer Acc65I is compatible with BsiWI and BsrGI) sites or by ligation-independent cloning.

```
EcoRI
GAATTC ....................insert...........................

KpnI
 G  T  G  G  S  G  G  S  G  L  N  D  I  F  E  A  Q  K  I  E
GGTACCGGAGGTTCCGGTGGTTCCGGTCTGAATGATATCTTTGAAGCTCAGAAGATTGAA

 W  H  E  G  R  T  K  H  H  H  H  H  H  *  *    XhoI
TGGCATGAAGGACGTACCAAGCACCACCATCACCATCACTAATGATCACTCGAG
```

Reference

Aricescu, A. R., Hon, W. C., Siebold, C., Lu, W., van der Merwe, P. A. & Jones, E. Y. (2006). *EMBO J* **25**, 701-712.