Journal of Structural Biology 172 (2010) 55–65

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Recent advances in the production of proteins in insect and mammalian cells for structural biology

Joanne E. Nettleship, René Assenberg [1], Jonathan M. Diprose, Nahid Rahman-Huq, Raymond J. Owens *

The Oxford Protein Production Facility and Division of Structural Biology, Henry Wellcome Building for Genomic Medicine, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

ARTICLE INFO

Article history:
Received 23 December 2009
Received in revised form 4 February 2010
Accepted 7 February 2010
Available online 11 February 2010

Keywords:
Baculovirus
High throughput
Transient expression
HEK293 cells

ABSTRACT

The production of proteins in sufficient quantity and of appropriate quality is an essential pre-requisite for structural studies. *Escherichia coli* remains the dominant expression system in structural biology with nearly 90% of the structures in the Protein Data Bank (PDB) derived from proteins produced in this bacterial host. However, many mammalian and eukaryotic viral proteins require post-translation modification for proper folding and/or are part of large multimeric complexes. Therefore expression in higher eukaryotic cell lines from both invertebrate and vertebrate is required to produce these proteins. Although these systems are generally more time-consuming and expensive to use than bacteria, there have been improvements in technology that have streamlined the processes involved. For example, the use of multi-host vectors, i.e., containing promoters for not only *E. coli* but also mammalian and baculovirus expression in insect cells, enables target genes to be evaluated in both bacterial and higher eukaryotic hosts from a single vector. Culturing cells in micro-plate format allows screening of large numbers of vectors in parallel and is amenable to automation. The development of large-scale transient expression in mammalian cells offers a way of rapidly producing proteins with relatively high throughput. Strategies for selenomethionine-labelling (important for obtaining phase information in crystallography) and controlling glycosylation (important for reducing the chemical heterogeneity of glycoproteins) have also been reported for higher eukaryotic cell expression systems.

1. Introduction

Over the last 10 years there have been major advances in the technology for recombinant protein production specifically for structural biology. Much of this has been led by structural genomics centres which have pioneered high throughput approaches to sample preparation including the use of laboratory automation to achieve parallel processing. *Escherichia coli* remains the dominant host for producing recombinant proteins as shown by an analysis of expression systems used for structures deposited in the Protein Data Bank (PDB) (Table 1). Thus 88% of the protein chains from structures for which expression system annotation is available were produced in *E. coli* compared to 9% for all eukaryotic hosts combined. Of these, baculovirus/insect cells accounted for 4%, whilst mammalian cells, including both stable cell lines and more recently transient expression, represented 2.4%. Although multi-construct approaches have increased the success rate for obtaining soluble protein from *E. coli* (Fogg et al., 2006; Graslund et al., 2008), there are proteins which are not amenable to bacterial expression, e.g., many membrane proteins, multi-protein complexes and cell surface or secreted glycoproteins (Aricescu et al., 2006b). It is estimated that as many as 50% of all human sequences may be glycosylated (Apweiler et al., 1999) representing a major challenge for structural biology in terms of producing soluble proteins. This has stimulated the use of higher eukaryotic cells for the production of proteins for structural studies, reflected in the steady increase in the number of structures deposited in the PDB of proteins produced using these systems (Fig. 1). Notable recent examples include the first reported structures of G-protein coupled receptors (Cherezov et al., 2007; Rosenbaum et al., 2007) and the ATP-gated ion channel P2X4 (Kawate et al., 2009), both produced by baculovirus infection of insect cells, and viral glycoproteins obtained from transient expression in mammalian cells (Bowden et al., 2008a,b). In this article we review the recent advances in insect and mammalian cell expression technology which have improved their ease of use, increasing both throughput and robustness of these techniques.

2. Baculovirus expression system

The baculovirus expression system is a well-established method for the production of recombinant proteins with the major

* Corresponding author.
E-mail address: ray@strubi.ox.ac.uk (R.J. Owens).
[1] Present address: Novartis Pharma AG/NIBR, Fabrikstrasse 16.2.16, CH-4056 Basel, Switzerland.

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.02.006

Table 1
Expression systems used for producing proteins for structural biology.

Expression system	No. chains	% of identified
Amoeba	26	0.05
Archaea	41	0.08
Avian	13	0.02
Bacteria	47162	88.08
Baculovirus	1879	3.51
Cell free	1435	2.68
Fungi/yeast	1364	2.55
Insect	296	0.55
Mammalian	1297	2.42
Plant	31	0.06
Unknown	33550	N/A
Total	87094	N/A

The number of chains deposited in the PDB by expression system, and as a percentage of the total number of chains with an identifiable expression system, as of December 2009. Information about the expression system was parsed from the set of PDB files available from: ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/pdb. Incomplete or inconsistent information was resolved to the best of the authors' ability or marked as unknown when unresolvable; the vast majority (33422 out of 33550) of the unknowns had no expression information. The system was assigned as baculovirus where the expression organism was indicated as *S. frugiperda* or *T. ni* in addition to when specifically indicated. Chains were counted rather than PDB entries as expression information is recorded by chain in the PDB.

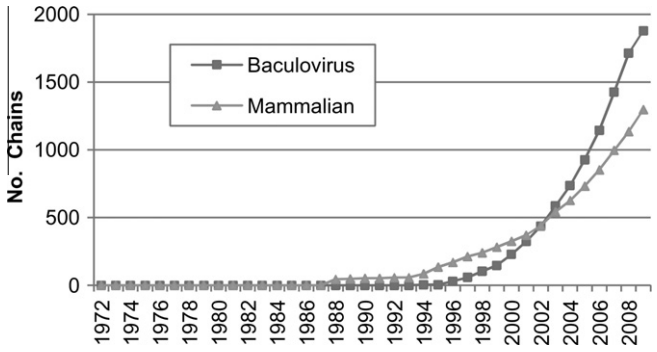

Fig. 1. Plot of the cumulative total number of chains deposited in the PDB whose expression system was identified as either baculovirus or mammalian by year of deposition. Expression data were parsed from PDB files as described for Table 1. The data for 2009 (and even some of 2008) are incomplete as some 3450 structures deposited over this period are yet to be released at the time of writing. Despite this, Mammalian-derived depositions have already risen ~16% from 2008 to 2009 (139–162 chains). Otherwise, the rate of deposition has been increasing for both systems in a trend that is similar to the overall rate of PDB depositions, though, as described by Levitt (2007), the growth is not exponential. For a more detailed analysis of the growth of the PDB as a whole without consideration of the expression system, see Levitt (2007).

advantage over *E. coli* that expressed proteins undergo post-translational modifications, e.g., phosphorylation, myristoylation and glycosylation. Unlike mammalian cells (see Section 3), the *N*-glycans attached to proteins expressed in insect cells are characterised by paucimannose-type structures (Harrison and Jarvis, 2006). The principal baculovirus used for recombinant protein expression is *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) with *Spodoptera frugiperda* 9 (*Sf*9) insect cells as the expression host. Traditionally, the construction of recombinant baculoviruses required time-consuming rounds of plaque purification to isolate the recombinant virus from a background of wild-type non-expressing viruses. However, the propagation of the baculovirus genome as a bacmid in *E. coli* has revolutionised the manipulation of the virus and opened the way for efficient and relatively high throughput virus generation (Section 2.1). Small-scale expression screening combined with easy-to-use disposable bioreactors has also greatly improved the efficiency of baculovirus expression technology (Sections 2.2 and 2.3).

2.1. Vectors and viruses

2.1.1. One-step virus production by homologous recombination

The construction of baculoviruses by transposition of the gene to be expressed into the BAC10 bacmid in *E. coli* (Luckow et al., 1993), commercialised as the Bac-to-Bac™ system (Invitrogen), is probably still the most widely used and enables the routine and rapid production of viruses. A wide variety of vectors are available for the system, including ones which have been engineered to enable insertion of genes using ligation-independent cloning methods, e.g., the Gateway® system (Abdulrahman et al., 2009). Alternative approaches to simplifying the generation of recombinant baculoviruses have focused on disabling the viral genome used for homologous recombination in insect cells as the means of generating recombinant viruses (Je et al., 2001; Zhao et al., 2003). In one example of this approach a version of the BAC10 bacmid was constructed in which the essential gene, ORF1629, had been inactivated by the insertion of the chloramphenicol acetyl transferase gene to produce BAC10:KO_{1629} (Zhao et al., 2003). Co-transfection of insect cells with linearized BAC10:KO_{1629} and any standard baculovirus transfer vector results in the generation of 100% recombinant viruses, removing the need for plaque purification. A similar approach has been described by Possee et al. (2008) with the added feature that the chitinase gene (*chiA*) has been deleted from the virus (see Section 2.1.2). The linearized bacmid and derivatives are commercially available as flashBAC™ (Oxford Expression Technologies Ltd.). Without the need to plaque purify viruses, both these systems can be readily automated for parallel virus construction (Possee et al., 2008). In addition, fewer steps are required compared to the Bac-to-Bac™ system thereby saving time in the virus construction process. As a note of caution, there is some evidence that all bacmid-derived baculovirus expression vectors undergo spontaneous deletion following repeated passaging of viruses but only following more than five rounds of amplification (Pijlman et al., 2003).

2.1.2. Multi-functional and co-expression vectors

A variety of transfer vectors are available for the construction of recombinant baculoviruses that encode resident fusion proteins which have been reported to improve protein expression, including maltose binding protein (Pengelley et al., 2006), glutathione S-transferase (Abdulrahman et al., 2009; Romier et al., 2006) and SUMO (Liu et al., 2008b). In addition, by incorporating promoter elements for expression in *E. coli* and/or mammalian cells, parallel screening of constructs in both bacterial, mammalian and insect cell systems can be carried out (Berrow et al., 2007; Chambers et al., 2004; Pengelley et al., 2006). This has the obvious benefit that switching between expression systems is easy and enables parallel screening to identify the best expression host. Using such a dual promoter vector, it has been reported that for a set of 62 human kinases, 29 were expressed as soluble proteins in *E. coli*, whereas 61 were obtained as soluble in insect cells (Chambers et al., 2004).

To address the structure of multi-protein complexes may require the co-expression of several components in the same host since combining individual components produced separately is often not possible due to poor levels of expression and/or solubility of the component proteins when expressed on their own. At its simplest level, co-infection of insect cells with recombinant viruses for each component can be used to identify suitable constructs and/or sub-complexes. To maximise co-infection and hence overall expression of the complex requires optimisation of the multiplicity of infection (MOI) for each virus. Alternatively, a single virus can be constructed expressing multiple genes by using a transfer vector consisting of each opening reading frame under the control of a separate promoter. Such dual expression vectors are available