High-throughput protein expression for the post-genomic era

Stephen P. Chambers

In the past, protein expression has been perceived as the principle bottleneck in protein characterization and structure determination. The challenge now is to rapidly express large numbers of genes in the search for new drug targets and therapeutic proteins encoded by the human genome. In this competitive environment, several high-throughput expression strategies for protein production are being used to industrialize the process of protein expression.

Stephen P. Chambers
Vertex Pharmaceuticals
130 Washington Street
Cambridge
MA 02139, USA
tel: +1 617 444 6122
fax: +1 617 444 6437
e-mail: chambers@vrtx.com

Dramatically illustrated in the sequencing of the human genome, high-throughput tools in biology have changed the nature and scale of experiments forever. With the identification of >35,000 genes in the human genome [1,2] there is an increasing demand for high-throughput expression of recombinant proteins.

An immediate challenge of the post-genomic era is to assign biological function to all the proteins encoded by the genome.

Initial computational analysis of the human genome using the basic local alignment tool (BLAST) [3] attempted to assign function to all the proteins predicted to be encoded. Using sequence similarity to proteins with known function, ~60% of the predicted proteins were functionally annotated. The remaining 40% could not be ascribed any function [1].

Understanding the biological function of these unknown proteins has emerged as the major task in the translation of the genome. Structure-based assignment of function is thought to provide advantages over annotation on the basis of sequence alone. Because sequence is generally less conserved than structure, structural homology can sometimes be a strong predictor of similar function, a relationship not necessarily identifiable from the sequences [4]. Driven by the need to produce many proteins in parallel and to identify the most suitable protein constructs for structural analysis, much of the development for high-throughput expression has been geared towards downstream efforts in protein structure determination [5].

In addition, high-throughput expression has been developed to generate protein for microarrays and antibodies for immunolocalization and functional analysis. For thorough characterization of any gene, its corresponding protein must be expressed and examined. This is the main driving force behind the development of a process that rapidly produces large numbers of functional proteins for both structural and enzymatic analysis.

From laboratory to factory

The application of parallel processing, miniaturization and automation to the sequencing of the human genome massively increased throughput [6]. Despite the difficulty imposed by protein complexity, these same techniques and principles are now being applied to protein production, in establishing new methodologies and tools for the high-throughput expression of proteins.

Historically, a variety of protein expression strategies were developed for the production of individual proteins. Relying on a degree of serendipity and the art of the experimentalist, these expression strategies evolved empirically. Unfortunately, this model does not lend itself to high-throughput and, to establish protein expression on the scale required for the genome analysis, robust generic methods need to be developed that simplify and stabilize the process to allow automation.

High-throughput cloning

The starting point for the translation of genetic information into protein is the cloning of the gene to be expressed into the appropriate vector. When attempting the

1359-6446/02/$ – see front matter PII: S1359-6446(02)02361-9

cloning of multiple genes as part of a high-throughput process, conventional cloning methods using restriction enzymes and DNA ligase to 'cut and paste' cannot be used. The combination of engineering unique cleavage sites for each gene that are compatible with the expression vector and the multiple purification steps required, severely limit the throughput of this approach.

Many of the disadvantages encountered in adapting conventional cloning to high-throughput have been eliminated with the development of Topo® cloning [7] (Invitrogen, Carlsbad, CA, USA). Using vaccinia virus DNA topoisomerase-I to covalently join polymerase chain reaction (PCR)-generated DNA fragments to a suitable vector provides a ligase-free method of cloning. This highly efficient alternative method of cloning requires only a 5 min room temperature incubation, with no special reaction conditions or additions and no post-PCR modifications (e.g. blunting or cutback), therefore making it amenable to automation [8].

The subsequent problem of sub-cloning into multiple vectors to access a variety of different expression systems is addressed in the section on expression vectors.

Expression systems

A variety of expression systems is available for the high-throughput production of recombinant proteins. *Escherichia coli* expression being the simplest, it has been the most readily adapted to a high-throughput format. There are several examples available in which high-throughput *E. coli* expression has been used to produce protein for functional [9,10] and structural [11] analysis.

Unfortunately, *E. coli* has all the advantages and disadvantages of simplicity. Although *E. coli* is a suitable starting point for the production of recombinant proteins of modest size and complexity, it is not well suited to the production of proteins with more-complex structures, particularly those requiring post-translational modifications for biological activity [12]. In addition, many recombinant proteins produced in *E. coli* accumulate intracellularly as insoluble, biologically inactive inclusion bodies [13]. Many of the limitations associated with *E. coli* and other prokaryotic expression systems can be alleviated using eukaryotic expression systems. Yeast, mammalian or insect cells can accomplish most post-translational modifications, which might be crucial for the biological integrity and function of recombinant proteins [14]. Because of the advantages associated with eukaryotic expression, several mammalian [7], yeast [15] and insect cell [16,17] expression systems have been successfully modified for high-throughput production and purification of proteins, despite the many technical challenges that are associated with them.

Cell-free expression systems

One of the simplest methods of protein synthesis involves *in vitro* protein expression using cellular lysates that contain all the necessary biochemical components for the transcription and translation of DNA templates. Many of the problems encountered when producing toxic and insoluble proteins in whole cells can be circumvented using cell-free expression systems.

In addition, in regard to high-throughput expression strategies, cell-free expression systems have distinct advantages over conventional *in vivo* synthesis. Specifically, proteins can be expressed directly from PCR-generated transcripts [18], thereby avoiding the need to clone genes into expression vectors. Eliminating the manipulation of cells in culture makes cell-free systems highly amenable to automation. Cell-free expression systems can be used to produce isotopically [19] and seleno-methionine [20] labelled protein suitable for structural examination.

In the past, the exploitation of cell-free expression technologies has been limited by the relatively low yields of protein produced by this method. Recent advances in cell-free expression significantly raised expression levels [21], making the technique viable for integration in high-throughput expression strategies. Using an *E. coli* cell-free system, PCR-generated DNA transcripts have been expressed and screened for protein production and solubility in a 96-well format [22].

This revival of interest in cell-free expression technologies has been marked by the increased use of prokaryote and eukaryote cell-free lysates for expression [23]. Further developments include the introduction of a rapid translation system (Roche Molecular Biochemicals, Indianapolis, IN, USA) that includes the reagents and instrumentation to produce the optimal reaction conditions for cell-free expression [24].

Expression vectors

In theory, the ideal expression strategy would use numerous expression vectors to access and exploit the individual and collective advantages provided by multiple expression systems. Unfortunately, the logistics of applying such a broad strategy when attempting high-throughput expression of several genes is prohibitive. In addition to the redundancy of effort, the process would involve the time consuming process of sub-cloning individual genes into numerous expression vectors. This traditional approach to sub-cloning is labor-intensive and not easily amenable to high-throughput. Although PCR has provided an efficient way of cloning genes, the random introduction of mutant sequences into genes amplified by PCR necessitates further time-consuming DNA sequence validation following any

PCR manipulation. This difficulty is compounded when sub-cloning into multiple vectors to access a variety of different expression systems.

Developments in recombination cloning have effectively eliminated traditional sub-cloning by replacing PCR and restriction enzyme–DNA ligase reactions with the site-specific transfer of genes from one vector to another via recombination. Commercialized recombination cloning methods such as Gateway™ cloning (Invitrogen) uses λ site-specific recombination reactions to perform integrations and excisions at specific (*att*) recombination sites [25]. Initially, the gene of interest is cloned into a specific entry vector between two distinct *att* recombination sites. The resulting entry vector, mixed *in vitro* with a suitable destination vector containing complementary *att* sites, undergoes recombination to produce an expression clone. There are multiple destination vectors available that provide access to several different expression vectors and systems. Gateway™ cloning technology is at the center of the FLEXGene repository established by the Institute of Proteomics at Harvard [26], which aims to create a comprehensive repository of production-ready cDNA clones. A similar system based on Cre-*lox*P recombination [27], marketed as the Creator System (Clontech, Palo Alto, CA, USA), is also being used by the Mammalian Gene Collection Project – a joint effort by the National Institutes of Health and the National Cancer Institute to provide full-length clones.

Although recombination cloning provides relatively easy access to several expression vectors compared with conventional sub-cloning, one defect arises in the transfer of recombination sites that accompany any gene on recombination. The resultant fusion can be potentially problematic in any subsequent structural or functional characterization.

To avoid the difficulties and problems, alternative multi-system vectors have been developed; pTriEx-1 vector [28] for expression in mammalian, *E. coli* and insect cells, and pBEV vector (Fig. 1) for expression in *E. coli* and insect cells. Multi-system vectors facilitate one-stop cloning and allow immediate expression in *E. coli*.

As with the recombination cloning systems, the multi-system vectors rely on recombination to access alternative expression systems. However, unlike the other recombination cloning systems described, the recombination sites are not incorporated into the coding region of the gene and are therefore not part of the expressed protein.

Depending on the multi-system vector used, recombination can take place in insect cells [29] or *E. coli* [30] to generate recombinant baculovirus. Once generated, recombinant baculoviruses can serve as gene entry vectors for the transfection of insect cells and the transduction of mammalian cells. Under the direction of insect or mammalian promoters, recombinant genes can be expressed in a wide range of insect and mammalian cells, respectively [31]. With the inclusion of a suitable marker in the viral vector, cell lines can also be selected for stable gene expression [32].

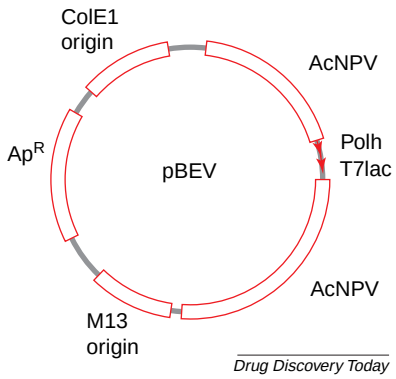

Figure 1. pBEV: a multi-system expression vector. The figure shows the gene for ampicillin resistance (Ap^R), the ColE1 origin of replication, the M13 origin of replication, the polyhedron (Polh) promoter, the T7lac promoter and the *Autographa californica* nuclear polyhedrosis virus (AcNPV) recombination sequence.

Protein expression

The throughput required for any kind of genome-scale protein expression requires an alternative to conventional cell cultivation. Using shake-flasks, even small shake-flasks, is impractical. Fortunately, deep-well blocks provide an alternative. Originally designed for growing bacterial cultures for DNA purification, deep-well blocks are readily adapted to cell growth for protein expression and are available in different formats (96-, 48- and 24-wells) and volumes (1, 2 and 5 ml per well). Covered with AirPore™ tape sheets (Qiagen, Valencia, CA, USA), which allows gaseous exchange and maintains sterility, these blocks perform much like shake-flasks in the growth of *E. coli* and insect cells. Although conventional shakers can be adapted to house deep-well blocks, a more efficient option is the HiGro™ shaker (GeneMachines, San Carlos, CA, USA) – a high-capacity incubating shaker specifically designed for a microtiter format, combining vortex mixing, gas flow and temperature control. The growth profiles of *E. coli* and insect cells using the HiGro Shaker is the same as that seen in shake flasks (Fig. 2).

After cultivation, cells are harvested by centrifugation and protein is extracted from the cells using chemical lysis or mechanical disruption. BugBuster (Novagen, Madison,

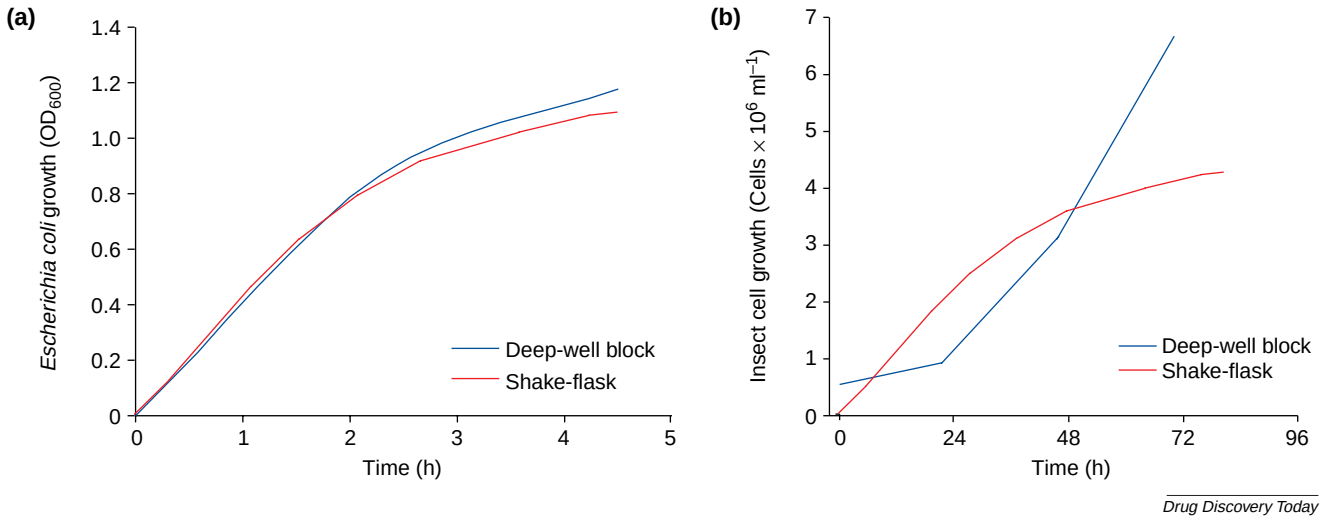

Figure 2. Growth-curves of **(a)** *Escherichia coli* and **(b)** insect cells, in shake-flask and deep-well block. Graphs show comparable growth over time between shake flask and deep-well block for both *E. coli* and insect cells.

WI, USA) – a specially formulated reagent that disrupts the cell wall and degrades nucleic acid, resulting in the liberation of soluble protein – is promoted specifically for high-throughput applications [33]. Alternatively, mechanical disruption using a deep-well cup-horn sonicator (Misonix Inc., Farmingdale, NY, USA) is ideal for cell disruption in deep-well plates.

Miniaturization of cell growth in this manner allows simultaneous protein production over a wide range of culture conditions (e.g. temperature, time, media, host) to screen for optimum expression. The development of the Micro-bioprocessor (Fluorometrix, Stow, MA, USA), which uses noninvasive sensor technology to monitor each well of a 24-well block for pH, optical density and dissolved oxygen [34], will further broaden the potential of miniaturized high-throughput expression by allowing greater control of the culture conditions.

Protein purification

Regardless of the expression vector or system used, high-throughput production requires the proteins to be tagged with a protein or peptide with properties that can be exploited in a common purification strategy. When working on a high-throughput platform, the ability to efficiently purify protein is of crucial importance and can only be achieved when the expressed proteins contain a common tag. The most common tag used is a sequence of six histidine residues (His-tag) that binds to immobilized divalent metal ions (Ni^{2+}), enabling rapid purification [35]. The His-tag also has the advantage of being easily identified by commercially available antibodies.

Larger tags are available, including *E. coli* maltose-binding protein (MBP) [36] and *Schistosoma japonicum* glutathione-S-transferase (GST) [37], both of which allow affinity purification of the fusion protein expressed. Fusions have been used successfully to overexpress many heterologous proteins in *E. coli* [38], and are often cited for their ability to improve solubility [39]. However, any advantage provided by a fusion is often negated by the unpredictability of fusion proteins, and by problems encountered downstream of protein expression [40].

Placement of tags at the N- or C-terminus of the protein also has its own set of advantages and disadvantages. The use of C-terminally tagged proteins has the advantage of eliminating any truncated proteins during purification. By contrast, N-terminal tags provide a common N-terminal codon sequence that is optimized for maximum expression [41], thereby furthering standardization during parallel expression of multiple proteins.

Tags provide a means of identifying and quantifying the amount of expressed protein. Some of the more common tags also allow a single-step purification to be automated using a liquid handling robot. Ni-nitrilotriacetic acid magnetic agarose beads with a high affinity for the His-tag have been developed for use in the magnetic separation and purification of His-tagged proteins in a 96-well format using a BioRobot™ (Qiagen) [42]. For greater binding capacity, SwellGel™ (Pierce, Rockford, IL, USA) Ni-chelating discs can be used for the single-step affinity purification of His-tagged proteins. SwellGel™ glutathione discs are also available, for the purification GST-tagged proteins. These products are available in a 96-well format and are readily

automated using a liquid handling robot. Depending on the expression level and the volumes used, between 1.5 μg and 6 mg of protein can be purified [43]. Although these levels are sufficient for a limited analysis determining expression levels, solubility and activity, they are not adequate for any extensive study. To meet the specific demands of high-throughput crystallography, large-scale protein purification has also been automated. Instrumentation to culture 96 × 65 ml of bacteria, with the robotics for centrifugation, sonication and purification, have been developed to yield, on average, 10 mg of purified protein per *E. coli* culture [11].

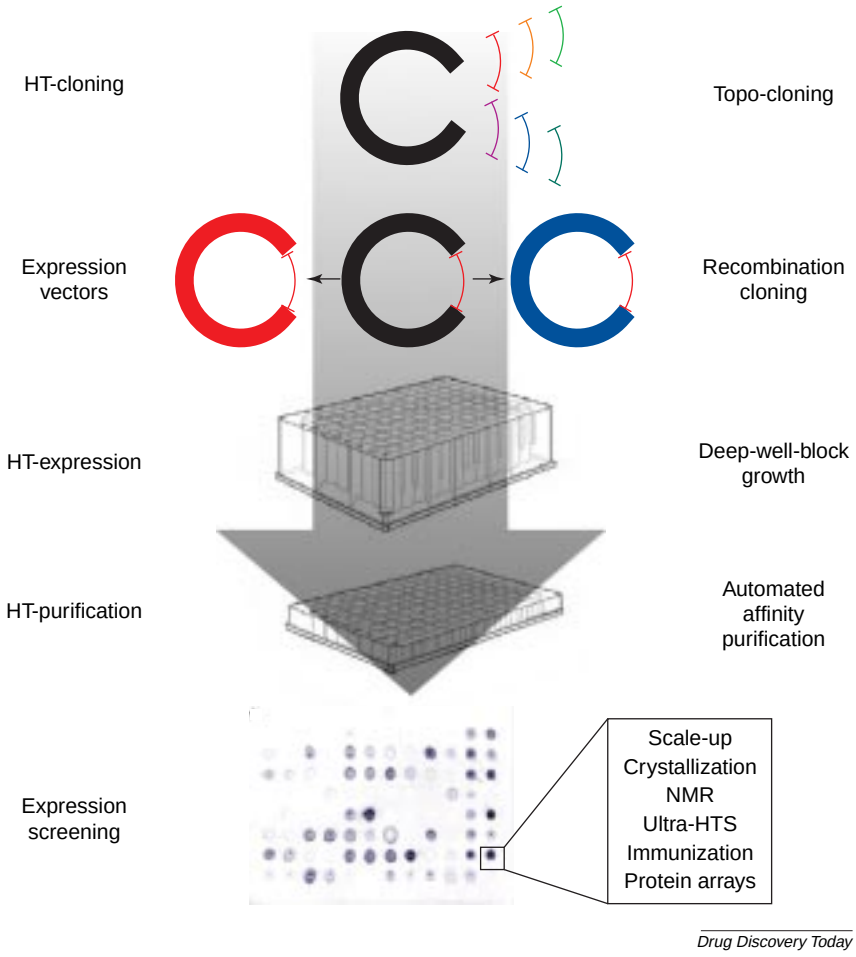

Figure 3. High-throughput expression: workflow and areas of development. The diagram identifies some of the tools used to advance high-throughput expression and cumulates with the identification of the end-users of this process. Abbreviations: HT, high-throughput; HTS, high-throughput screening; NMR, nuclear magnetic resonance.

Screening for expression

The application of automated, liquid-dispensing workstations for optimizing crystallization conditions [44], and the development of micro-crystallization [45], have significantly reduced the amount of protein required for crystallography. However, the amount of concentrated, homogeneous, monodispersed and soluble protein required for crystallization is still significant, and presents a challenge for any of the high-throughput expression strategies described.

Alternatively, high-throughput expression can be used to screen many proteins, identifying those that are soluble and/or overexpressed [46]. Protein solubility in aqueous buffers is often taken as evidence of native-like folding, essential for any structural or functional characterization. By contrast, soluble overexpression is often a good predictor of success in crystallography and nuclear magnetic resonance (NMR).

For the more intractable proteins, it has been shown that protein solubility can be dramatically increased by amino acid substitutions at certain positions [47,48], and by the deletion of certain portions of the protein [49]. In this manner, high-throughput expression has been used to screen and select proteins for solubility from a pool of deletions and mutants. Similarly, proteins fused to the N-terminus of green fluorescent protein (GFP) can also be screened for solubility using the *in vivo* fluorescence from GFP as an indicator of solubility [50]. Other reporter genes, including chloroamphenicol acetyl transferase [51] and β-galactosidase [52], have also been used to screen and detect soluble expressed protein. The disadvantage with a reporter system is that, as a fusion, it might affect the solubility of the protein of interest. This problem has been circumvented by coupling the upregulation of genes responsive to translational misfolding to β-galactosidase activity and screening for lack of activity [53]. By using these methods, many proteins can be screened for both expression and solubility and, in this way, proteins can be rapidly triaged before production (Fig. 3). Once the protein has been identified as expressed or soluble (preferably both) it can be produced at the scale required.

Conclusions

Many of the high-throughput methods described here have been developed as pilot studies – proposed solutions

to the proteomic challenge of large-scale protein function and biochemical analysis. These studies have been performed on a relatively small-scale when compared to the size of the human genome, and have been very selective with respect to their choice of proteins or genomes studied. Purposely working on only the most tractable proteins to develop the processes over-simplifies the problems that will be encountered when expressing the full spectrum of proteins encoded by the human genome.

How effective any one of these high-throughput methods will be at resolving the many problems that will be encountered when expressing the genome is uncertain. What is certain is that high-throughput techniques are changing the way protein expression is performed. The basic questions asked of protein expression are: what expression system; what vector; what protein construct; and which growth conditions? These questions and others can be answered by operating protein expression as a process and by applying high-throughput techniques.

It is unlikely that high-throughput expression, in its current form, is going to meet the current need for the copious amounts of protein that are required for high-throughput screening (HTS) and structural determination. Even with the advent of ultra-HTS and micro-crystallization, the amounts of protein required for drug discovery are significant and a challenge for any of the high-throughput expression systems described.

The most significant change to accompany high-throughput expression results from the massive increase in the number of proteins that can now be processed in parallel. The economies of scale and efficiencies produced using high-throughput expression produce recombinant proteins as a commodity, which can be delivered as a product rather than a service. This fundamental change is vital if the promises of the genome are to be realized.

References

1 Venter, J.C. *et al*. (2001) The sequence of the human genome. *Science* 291, 1304–1351
2 Lander, E.S. *et al*. (2001) Initial sequence and analysis of the human genome. *Nature* 409, 860–921
3 Altschul, S.F. *et al*. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
4 Shapiro, L. and Harris, T. (2000) Finding function through structural genomics. *Curr. Opin. Biotechnol.* 11, 31–35
5 Stevens, R.C. (2000) Design of high-throughput methods of protein production for structural biology. *Struct. Fold. Design* 8, R177–R185
6 Hawkins, T.L. *et al*. (1997) DNA sequencing: a magnetic attraction to high-throughput genomics. *Science* 276, 1887–1889
7 Heyman, J.A. *et al*. (1999) Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* 9, 383–392
8 Nasoff, M. *et al*. (2000) High-throughput expression of fusion proteins. *Methods Enzymol.* 328, 515–529
9 Bussow, K. *et al*. (2000) A human cDNA library for high-throughput protein expression screening. *Genomics* 65, 1–8
10 Larsson, M. *et al*. (2000) High-throughput protein expression of cDNA products as a tool in functional genomics. *J. Biotechnol.* 80, 143–157
11 Lesley, S.A. (2001) High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expres. Purif.* 22, 159–164
12 Rademacher, T.W. *et al.* (1988) Glycobiology. *Annu. Rev. Biochem*. 57, 785–838
13 Williams, D.C. *et al.* (1982) Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science* 215, 687–689
14 Vlak, J.M. and Keus, R.J.A. (1990) Baculovirus expression vector system for production of viral vaccines. *Adv. Biotechnol. Processes* 14, 91–128
15 Schuster, M. *et al*. (2000) Protein expression strategies for identification of novel targets. *J. Biomol. Screen*. 5, 89–97
16 Albala, J.S. *et al*. (2000) From genes to proteins: high-throughput expression and purification of the human proteome. *J. Cell. Biochem*. 80, 187–191
17 Coleman, T.A. *et al*. (1997) Production and purification of novel secreted human proteins. *Gene* 190, 163–171
18 Martemyanov, K.A. *et al.* (1997) Direct expression of PCR products in a cell-free transcription/translation system: synthesis of antibacterial peptide cecropin. *FEBS Lett.* 414, 268–270
19 Kim, D-M. *et al*. (1996) A highly efficient cell-free protein synthesis system from *E. coli*. *Eur. J. Biochem*. 239, 881–996
20 Kigawa, T. *et al*. (2002) Selenomethionine incorporation in to a protein by cell-free synthesis. *J. Struct. Funct. Genomics* 2, 29–35
21 Kigawa, T. *et al*. (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett*. 442, 15–19
22 Yokayama, S. *et al*. (2000) Structural genomics in Japan. *Nat. Struct. Biol.* 7 (Suppl.), 943–945
23 Madin, K. *et al.* (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 559–564
24 Martin, G.A. *et al*. (2001) High-yield, *in-vitro* protein expression using a continuous-exchange, coupled transcription/translation system. *BioTechniques* 31, 948–953
25 Hartley, J.L. *et al*. (2000) DNA cloning using *in-vitro* site-specific recombination. *Genome Res.* 10, 1788–1795
26 Brizuela, L. *et al*. (2001) FLEXGene repository: from sequenced genes to gene repositories for high-throughput functional biology and proteomics. *Mol. Biochem. Parasitol.* 118, 155–165
27 Liu, Q. *et al*. (1989) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* 8, 1300–1309
28 Novy, R. *et al*. (1999) pTriEx multisystem vector for protein expression in *E. coli*, mammlian and insect cells. *InNovations* 10, 1–5
29 Smith. G.E. *et al*. (1983) Production of human β-interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell Biol.* 3, 2156–2165
30 Luckow, V. *et al*. (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus geneome propagated in *E. coli*. *J. Virol.* 67, 4566–4579
31 Condreay, J.P. *et al*. (1999) Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc. Natl. Acad. Sci. U. S. A.* 96, 127–132
32 Merrihew, R.V. *et al*. (2001) Chromosomal integration of transduced recombinant baculovirus in mammalian cells. *J. Virol.* 75, 903–909
33 Grabski, A. *et al*. (1990) BugBuster and benzonase: the clear solution to simple, efficient extraction of *E. coli* proteins. *InNovations* 10, 17–19
34 Kostov, Y. *et al*. (2000) Low-cost microbioreactor for high-throughput bioprocessing. *Biotechnol. Bioeng.* 72, 346–352
35 Hochuli, E. *et al*. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent. *Biotechnology* 6, 1321–1325

36 Maina, C.V. *et al.* (1988) An *E. coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* 74, 365–373

37 Smith, D.B. and Johnson, K.S. (1988) Single-step purification of peptidase expressed in *E. coli* as fusions with glutathione S-transferase. *Gene* 67, 31–40

38 Pryor, K.D. and Leiting, B. (1997) High-level expression of soluble protein in *E. coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expres. Purif.* 10, 309–319

39 Kapust, R.B and Waugh, D.S. (1999) *E. coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668–1674

40 Riggs, P.D. (1992) Expression and purification of maltose-binding protein fusions. In *Current Protocols in Molecular Biology*, Greene/Wiley Interscience, NY, USA

41 Chen, G.T. and Inouye, M. (1990) Suppression of the negative effect of minor arginine codons on gene expression: preferential usage of minor codon within the first 25 codons of the *Escherichia coli* genes. *Nucl. Acid Res.* 18, 1465–1473

42 Lanio, T. *et al.* (2000) Automated purification of His_6-tagged proteins allows exhaustive screening of libraries generated by random mutagenesis. *BioTechniques* 29, 338–342

43 Romer, U. *et al.* (2001) Significantly higher yields from automated protein purification procedures. *QIAGEN News* 5, 12

44 Stewart, P.D.S. and Baldock, P.F.M. (1999) Practical experimental design techniques for automatic and manual protein crystallization. *J. Cryst. Growth* 196, 665–673

45 Pebay-Peyroula, E. *et al.* (1997) X-ray structure of bacteriorhodopsin at 2.5 Å from micro-crystals grown in lipidic cubic phases. *Science* 277, 1676–1681

46 Knaust, R.K.C. and Nordlund, P. (2001) Screening for soluble expression of recombinant proteins in a 96-well format. *Anal. Biochem.* 297, 79–85

47 Wetzel, R. *et al.* (1991) Mutations in human interferon γ affecting inclusion body formation identified by a general immunochemical screen. *Bio/Technology* 9, 731–737

48 Jenkins, T.M. *et al.* (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6057–6061

49 Barwell, J.A. *et al.* (1995) Overexpression, purification, and crystallization of the DNA binding and dimerization domains of the Epstein-Barr virus nuclear antigen 1. *J. Biol. Chem.* 270, 20556–20559

50 Waldo, G.S. *et al.* (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695

51 Maxwell, K.L. *et al.* (1999) A simple *in-vivo* assay for increased protein solubility. *Protein Sci.* 8, 1908–1911

52 Wigley, W.C. *et al.* (2001) Protein solubility and folding monitored *in vivo* by structural complementation of a genetic marker protein. *Nat. Biotechnol.* 19, 1359–1367

53 Lesley, S.A. *et al.* (2002) Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Eng.* 15, 153–160