BioTechnologia vol. 92(2) • pp. 119-131 • 2011
Journal of Biotechnology, Computational Biology and Bionanotechnology PERSPECTIVES

High throughput protein production

ALEKSANDER TWORAK, JAN PODKOWIŃSKI, MAREK FIGLEROWICZ*

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

* Corresponding author: marekf@ibch.poznan.pl

Abstract

Research in the relationship between the architecture and function of proteins aims to understand the mechanism of proteins folding and their activity at the atomic level. This knowledge leads to a progress in new technologies that allow simultaneous production of numerous proteins on a large scale – high throughput protein production (HTPP) methods. The HTPP methods allow parallel processing of multiple samples and are also important for the determination of optimal protein production procedure, where usually combinations of different conditions have to be tested. This article describes available cloning, expression, and purification approaches that may be used in a high throughput and parallel manner. Additionally, the implementation of automation facilities is briefly outlined.

Key words: protein, production, high throughput, structural genomics

Introduction

Over the last two decades an outstanding progress has been made in DNA sequencing methods. It took the Human Genome Project (HGP) consortium (established in 1990) 13 years to reveal 99% of the human genome sequence with 99.99% accuracy (Collins et al., 2003). By contrast, within the 1000 Genomes Project initiated in December 2008, 2500 human genomes were sequenced by the end of 2010 (Durbin et al., 2010). This incredible throughput increase was made possible due to the next-generation DNA sequencing technologies which became widely available a few years ago. A great impact of these technologies is even clearer when we look at a publicly available genome and the gene database (http://www.ncbi.nlm.nih.gov). As of January 2011, the NCBI Genome Resource lists more than a thousand eukaryotic genome sequencing projects and another thousand of complete microbial genomes, whilst GenBank stores over 100 million sequences from around 250 000 organisms. This includes complete sets of protein encoding genes for thousands of species. However, even such a massive amount of sequence data is not sufficient to deduce the complete protein composition of cells or the function of each single protein (Venter et al., 2001). And as our knowledge about complex biological systems depends on understanding the structure, function and interactions between many different molecules in the cell and beyond it, this knowledge needs to be gained at the protein level as well.

A systematic large scale proteomics study requires efficient methods of synthesis and purification of hundreds proteins in parallel. This high throughput protein production (HTPP) is still a few steps behind DNA technologies, which is nicely illustrated by the difference between the volumes of nucleotide sequence resources mentioned above and the total number of protein structures deposited in the Protein Data Bank (PDB, http://www.rcsb.org/pdb) – which roughly exceeds 70 000. The source of this difference is the nature of protein molecules, which – in contrast to nucleic acids – differ from one another in their physical and chemical properties. DNA molecules consist of a unified hydrophilic, negatively charged backbone and just four alternative components of the same nature: nucleotide bases. In contrast, proteins are made from 20 different amino acids that posses a side chain of different properties: hydrophilic or hydrophobic, charged or uncharged, polar or non polar. In the cell they can further undergo numerous post-translational modifications. Therefore, two proteins may be totally dissimilar in size, charge, stability, solubility

and in many other ways. This heterogeneity enables proteins to perform different biological functions, but impedes the application of unified handling procedures.

A number of different technologies, including numerous techniques of cloning, expression, and purification have been developed as an answer to the demand for high throughput protein production. All of them share a common feature: they facilitate an automated, parallel operation on multiple probes, in relatively small volumes. These three factors – automatization, parallel processing and small volume of samples are particularly important, as they save time as well as resources, and enable quick screening of the optimal conditions for each target protein and for each step of the production procedure (see Fig. 1).

Cloning systems

Preparation of several different constructs for each gene of interest is usually required in order to test various combinations and factors that are known to influence protein production efficiency. A generation of multiple expression vectors relies on the availability of an efficient and high throughput amenable cloning system.

In a traditional cloning procedure, cDNA that encodes a particular protein is PCR amplified with the use of a specific set of primers that provide restriction sites on its both ends. Those sites need to be picked carefully, so that they do not appear in the coding sequence and are compatible with an expression vector. Both the PCR product and the vector are digested with restriction endonucleases and a final construct is assembled with the use of DNA ligase. This procedure provides a great flexibility as many commercial expression vectors are available but, at the same time, is quite labor-expensive and time-consuming. For the latter reasons, its application in high throughput technologies is not recommended. Instead, more suitable cDNA cloning methods have recently been developed.

The Flexi Vector System (Promega) is based on directional cloning of protein coding sequence flanked by two rare restriction sites (SgfI and PmeI) into compatible vector. The chance that those restriction sites occur in the target sequence is very low, e.g. only about 1% of all human genes possess at least one such site. Therefore, almost every coding sequence may be cloned into a variety of Flexi Vectors and easily transferred between them. All the vectors carry the lethal barnase gene, which is replaced by the target sequence and acts as a selection factor for proper constructs. Because this system always utilizes the same set of enzymes to produce many insert-vector combinations, it is easily adaptable to high throughput manner (Blommel et al., 2009). Although the Flexi Vector family is limited, it nevertheless provides a high level of flexibility with regard to expression hosts or protein tags. Flexi Vector was shown to perform equally well as the most popular high throughput technology: Gateway system (Blommel et al., 2006).

All alternative cloning systems substitute traditional molecular biology tools such as restriction enzymes and ligase, by either some form of a recombination event or generation and subsequent annealing of complementary single-stranded overhangs. All of them are also designed to be fully independent of the input sequence.

Most widely used techniques are based on the site-specific recombination (SSR). This type of recombination occurs between strictly defined nucleotide sequences that are much longer than the typical restriction site. SSR is characterized by high precision and specificity. There are two alternative SSR cloning systems: Gateway (Invitrogen) and Creator (Clontech). Both systems are derived from different natural sources (bacteriophage lambda and bacteriophage P1, respectively) and differ in technical details due to their origin (Hartley et al., 2000; Liu et al., 1998). Gateway, the most popular technology, depends on four types of recombination attachment sites (*att* sites, about 25 nucleotides long; see Fig. 2 for details). The recombination occurs between a donor (usually a vector or PCR product) that contains the sequence of interest flanked by *att* sites of one type and an acceptor (destination vector) that contains two *att* sites of other type surrounding a lethal *ccdB* gene (Bernard, 1996). As a result, sequences flanked by *att* sites are exchanged between the donor and the acceptor. All undesirable recombination products are eliminated through *ccdB*-dependent negative selection. The Gateway system is based on the idea of a universal entry clone construction. The clone has a verified sequence and can be further used for transferring the gene of interest to any compatible expression vector. An easy and flexible way of entry clone production presumes *att* site-mediated recombination between cDNA and plasmid. The *att* sites can be introduced into cDNA during PCR amplification. The specially designed Gateway site-specific combination is rapid and accurate. It maintains proper orienta-

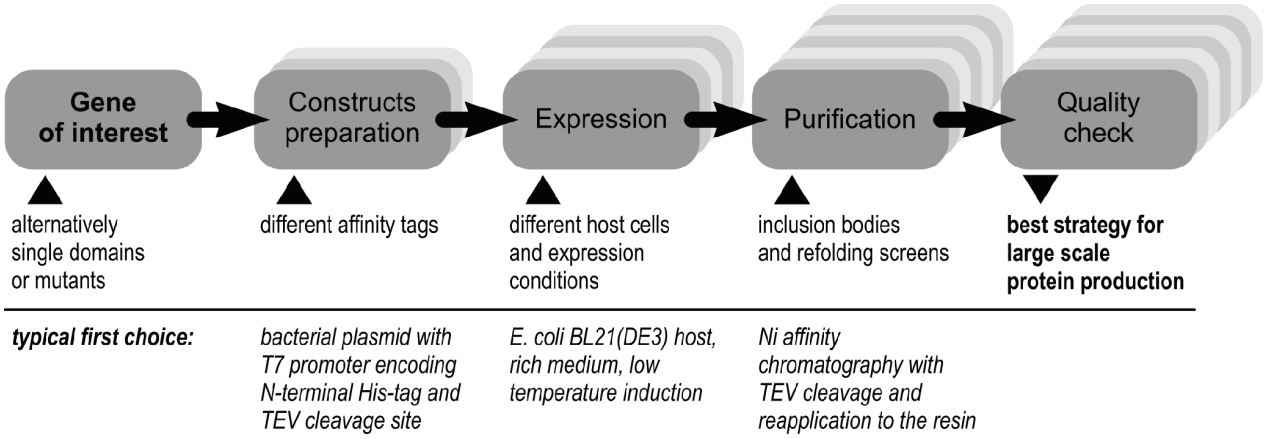

Fig. 1. Overview of the steps involved in a high-throughput protein production, highlighting the screening possibilities and typical first choice parameters

tion and a reading frame, so there is no need to resequence the final expression vector. SSR is also relatively insensitive to DNA concentration which overall makes it very suitable for high throughput approach. Numerous Gateway compatible expression vectors are available. Some have been designed for various protein synthesis and purification strategies, whereas others may be applied for *in vivo* protein analysis. Some potential disadvantages of the Gateway system are that in many cases *att* sites are incorporated into the coding sequence. This may influence protein structure or solubility. In addition, the system tends to lose its efficiency for larger inserts (over 3 kb).

Another type of cloning system: ligation-independent cloning (LIC) is based on the generation of 12-15 nucleotide (nt) long complementary overhangs in PCR product and a destination vector (Fig. 2). In a standard procedure, the vector is linearized by a restriction enzyme and the insert is synthesized by PCR. The corresponding 12–15 nt ends of both vector and insert are identical. In order to generate complementary single-stranded overhangs, the identical 12-15 nt ends need to be composed of only three types of nucleotides. The fourth missing nucleotide constitutes the first base pair in both ends. Deoxyribonucleotide, complementary to the missing one, is added to LIC reaction mixture along with T4 DNA polymerase. The enzyme generates overhangs through its 3′ to 5′ exonuclease activity but stops at the point where it can incorporate the dNTP through its polymerase activity. Approximately 12-15 nucleotide overhangs that are generated anneal strong enough to allow transformation of *E. coli* without prior ligation. Unligated ends are joined inside the bacterial cells by repair enzymes. The whole procedure does not require specialized vectors, the reagents are relatively inexpensive and only small, but high quality, amounts of the vector and insert DNA are needed. The greatest limitation is that the sequences of the annealing regions must lack one nucleotide. This means that the LIC vector has a specially designed sequence which is usually located in the reading frame and adds extra amino acids to the protein. In commercially available LIC-compatible vectors (Novagen), the number of redundant amino acids is minimized to three at each terminus.

An In-Fusion cloning system from Clontech functions according to the same principles as LIC (Fig. 2). In the case of In-Fusion system, about 15 nucleotide long complementary overhangs are generated by poxvirus DNA polymerase which, similar to T4 polymerase, has 3′ to 5′ exonuclease activity (Berrow et al., 2009). This enzyme, however, unlike the T4 polymerase, does not require any specific sequences to be stopped. Poxvirus DNA polymerase progressively removes nucleotides from the 3′ end of corresponding linear dsDNA, as long as the complementary regions can spontaneously hybridize. If an insert-vector junction contains a nick, gap or short overhang, the polymerase is stopped because it has lower affinity for any of those structures than for dsDNA ends. The main advantage of the In-Fusion system is that it does not depend on any specific sequence motifs. In consequence, additional amino acid codons are not introduced to the target protein coding sequence during vec-

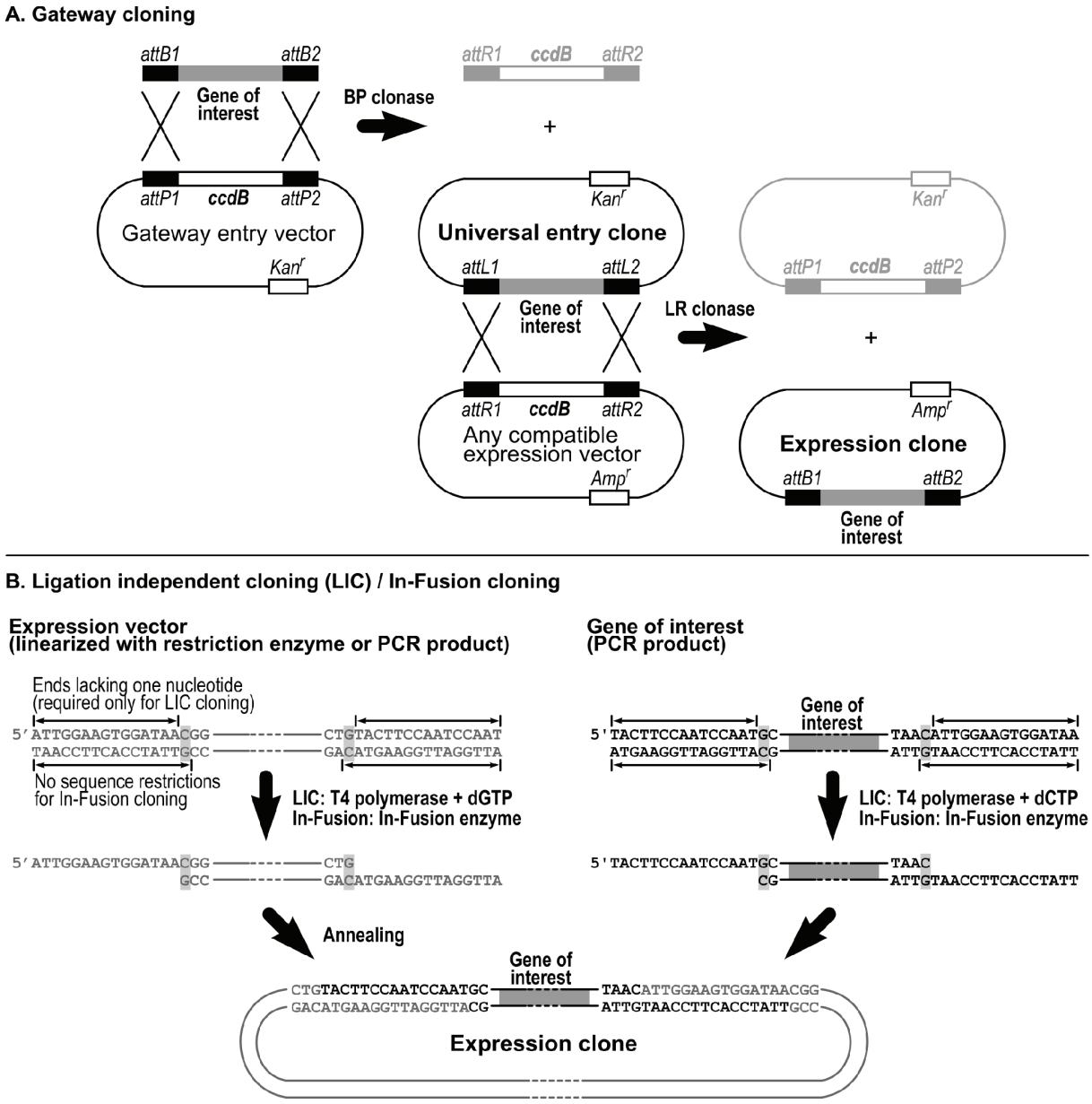

Fig. 2. Comparison of Gateway, Ligation independent (LIC) and In-Fusion cloning systems used for the construction of clones that express the gene of interest. (A) Schematic representation of the standard Gateway reactions that make use of four types of recombination sites: *attB*, *attP*, *attL* and *attR*. The first reaction (with BP clonase) facilitates the recombination of a gene of interest flanked by *attB* sites and an entry vector which contains *attP* sites. As a result, a universal entry clone with *attL* sites is generated. The gene of interest may be also cloned into an entry vector by Topo cloning (not shown). The entry clone is used for recombination with various compatible expression vectors containing *attR* sites (reaction catalysed by LR clonase). The vectors carry toxic *ccdB* gene, and kanamycin (*Kan*^{r}) or ampicillin (*Amp*^{r}) resistance gene that enable easy selection of proper constructs. Gateway cloning is directional because sequences of left and right cloning sites (like *attB1* and *attB2*) are not identical. (B) Schematic of Ligation independent (LIC) and In-Fusion cloning reactions. Both systems are based on a generation of approximately 15 nucleotides long complementary overhangs in PCR product and destination vector. The destination vector needs to be linearized by either restriction enzyme cleavage or PCR amplification. Both T4 polymerase (LIC) and In-Fusion enzyme generate 5′ overhangs utilizing their 3′ to 5′ exonuclease activity. T4 polymerase reaction is separately conducted for an insert and a vector in the presence of one dNTP: the enzyme is stopped at the point where its 5′ to 3′ polymerase activity counteracts the exonuclease activity. Therefore, for LIC cloning both vector and insert ends need to lack one nucleotide. In the case of In-Fusion enzyme, both the insert and the vector are mixed and treated in a single reaction: the exonuclease activity is interrupted as soon as complementary regions hybridize spontaneously. In this case, there are no restrictions in the nucleotide composition of the complementary ends, making this system applicable to a broader range of expression vectors. In both cloning systems, covalent bond formation at the insert-vector junctions occurs within the bacterial cell

tor construction. Moreover, the system can be used to join several pieces of dsDNA in a single reaction. All these features of In-Fusion system provide large flexibility in the expression vector design. Any sequence encoding an affinity tag or protease cleavage site may be combined with the target protein coding sequence (Benoit et al., 2006). However, opposite to the Gateway system, in the In-Fusion system protein coding sequences cannot be transferred between different vectors in a single reaction.

Interestingly, annealing of about 15 nucleotides long complementary overhangs may also be achieved without the use of any specific enzyme. This is possible because in a typical PCR reaction fully double stranded products are accompanied by the products that are not fully extended (Olsen and Eckstein, 1989). The Polymerase Incomplete Primer Extension (PIPE) cloning method is based on the use of PCR amplified vector and insert, both containing complimentary ends, partially single-stranded, due to specific PCR conditions (Klock and Lesley, 2009).

All of the described systems have been found useful for some particular strategies of high throughput cloning. Gateway, the most widely used cloning system, was implemented in numerous automated pipelines for the production of a large number of recombinant proteins of different origin, including *Arabidopsis thaliana* (Gong et al., 2004), humans (Lamesch et al., 2007; Nagase et al., 2008) and *Caenorhabditis elegans* ORFeome (Lamesch et al., 2004; Luan et al., 2004). In-Fusion technology was chosen by the Oxford Protein Production Facility that is focused on high throughput structure determination of proteins, many of which are problematic to synthesize and crystallize. Therefore the Oxford Protein Production Facility uses various In-Fusion constructed vectors, each one suited for protein synthesis in multiple hosts (*E. coli*, mammalian and insect cell lines) without any undesirable amino acids added to the product (Berrow et al., 2007). LIC cloning was found to be most cost-effective (compared to Gateway and In-Fusion) by the Structural Proteomics In Europe (SPINE) consortium, where various preselected targets were cloned and screened for expression (Alzari et al., 2006). Finally, traditional cloning (with two defined restriction enzymes) was adapted by the Protein Structure Factory (PSF) because of the relatively low costs and minimal impact on the coding sequence: addition of just two amino acids (Sievert et al., 2008).

Affinity tags

Affinity tags are the universal tools that can be fused to each recombinant protein to enable its quick and simple purification (Arnau et al., 2006). The tags enable simultaneous purification of highly dissimilar proteins with the use of one standardized procedure, which is essential for high throughput technologies. Although affinity tags were originally developed for purification purposes only, some of them have proved to enhance protein synthesis, folding, solubility, prevent proteolysis or protect the antigenicity of the fusion protein (Rajan et al., 1998; Kou et al., 2007; Hammarstrom et al., 2002; Mayer et al., 2004). In general, they fall into two major classes, depending on the type of the affinity target: they bind either to other peptides or small ligands linked to a solid support (see Table 1 for details).

The first group includes quite short tags that bind to a large affinity partner. The most commonly used examples are: calmodulin binding peptide (CBP), Strep II, S-tag, c-myc and FLAG-tag. The last two are classified as epitope tags as they bind to immobilized monoclonal antibodies. In general, all of these tags are quite expensive to use because of relatively high price and low capacity of the resin (immobilized proteins) but offer high degree of specificity for their binding partners (Waugh, 2005). And yet, their suitability for large scale HTPP is limited.

Three, by far the most popular, affinity tags, namely: glutathione S-transferase (GST), maltose binding protein (MBP), and polyhistidine tag (His-tag) belong to the second group (Derewenda, 2004). Two of them: GST and MBP are large peptides that bind to small molecules: glutathione and maltose, respectively. Both may increase the protein solubility in *E. coli* and provide an optimum context for translation initiation, in addition to the role of an affinity tag. They are also relatively cost-effective in terms of use but present some disadvantages of which the most important is high metabolic burden. Additionally, GST naturally forms dimers in the solution and may also aggregate due to the formation of disulfide bonds between the highly exposed cysteine residues present at the surface of each monomer (Kaplan et al., 1997). Those features may complicate affinity purification, especially in the case of oligomeric proteins. Furthermore, GST exhibits slow binding kinetics to glutathione sepharose resin, which makes loading of cell extracts very time consuming, especially at a large scale production stage.

Table 1. Comparison of commonly used affinity tags

Tag	Size	Affinity target	Features	Capacity (mg/ml)	RC[a]	Cost/1g[b]
CBP	26aa (4kDa)	Calmodulin	low metabolic burden high specificity unsuitable for expression in eukaryotic cells	2	3	$415
Strep II	8aa (< 2kDa)	Streptavidin variant	low metabolic burden high specificity	1.5	5	$7 500
S-tag	15aa (< 2kDa)	S fragment of RNase A	low metabolic burden high specificity	0.5	3	$9 000
FLAG	8aa (< 2kDa)	Monoclonal antibody	low metabolic burden high specificity	0.6	3	$56 000
GST	211aa (26kDa)	Glutathione	efficient translation initiation may enhance solubility high metabolic burden	8-10	5	$330
MBP	396aa (40kDa)	Amylose	efficient translation initiation may enhance solubility high metabolic burden	3	5	$260
His-tag	6aa (< 1kDa)	Transition metal ions	low metabolic burden purification possible under native and denaturing conditions background from endogenous proteins with multiple His residues	up to 40	5	$30

[a] Number of recommended regeneration cycles; [b] Cost of the resin calculated for purification of 1 g of recombinant protein based on the nominal binding capacity and the number of recommended regeneration cycles (Fong et al., 2010)

The last affinity tag is typically composed of 6 histidine residues that bind to immobilized transition metals ions, e.g. nickel. It combines the advantages of small size (low metabolic burden) with high capacity and low cost of the resin. Ni(II)-nitrilotriacetic acid (Ni-NTA), typical binding matrix of His-tagged protein capture, is capable of withstanding multiple regeneration cycles under stringent sanitizing conditions. The purification process may be done under native as well as strongly denaturing conditions required to solubilize inclusion bodies. All these features make His-tag the most widely used affinity tag for purifying recombinant proteins for biochemical and structural studies.

Although extremely practical, in some cases, affinity tags were observed to trigger negative effects. They included: changes in protein conformation, reduction of protein yields, alteration or total inhibition of enzyme activity and even toxicity (Chant et al., 2005; Goel et al., 2000; Fonda et al. 2002). The presence of an affinity tag, especially a large one, may be also unwanted in some particular applications such as structural studies or clinical use (Smyth et al., 2003). Thus the ability to remove the tag as soon as it has served its purpose to obtain a homogeneous protein product of native size and sequence is often desirable. The most common approach is to include protease cleavage recognition site at the junction between the tag and protein sequence. In this way a slightly longer but removable tag is created and additional purification strategy becomes available after the affinity chromatography the tag is cleaved off by a site-specific endopeptidase and subsequently another round of purification, with the same resin, is conducted. Any proteins that bind nonspecifically along with the released tag and the protease (if it also possesses the tag – which is a common case) remain on the resin while purified protein flows through (Arnau et al., 2006).

Many different site-specific endopeptidases are available, the most popular examples are: mammalian gastric protease enterokinase, two enzymes of mammalian blood clotting cascade: Factor Xa and Thrombin and two viral proteases: tobacco etch virus (TEV) protease and human rhinovirus (HRV) protease – see Table 2 for details (Charlton and Zachariou, 2011). Each enzyme recognizes a 4-8 amino acids long sequence that is highly unlikely to be found within the protein of interest. Notably, all of them make the cleavage directly after the recogni-

Table 2. Comparison of proteases commonly used for the removal of fusion affinity tags

Protease	Recognition site	Features
Enterokinase (serine endopeptidase)	Asp-Asp-Asp-Asp-Lys↓	reduced cleavage before Pro, site naturally present in FLAG-tag
Factor Xa (serine endopeptidase)	Ile-Glu/Asp-Gly-Arg↓	unlikely to cleave before: Pro, Arg
Thrombin (serine endopeptidase)	Leu-Val-Pro-Arg↓Gly-Ser	unlikely to cleave before: Pro
TEV (cysteine endopeptidase)	Glu-Asn-Leu-Tyr-Phe-Gln↓Gly	cleaves with various efficiencies before all other amino acids except Pro, high specificity, efficient at low temperatures
HRV (e.g. PreScission) (cysteine endopeptidase)	Leu-Glu-Val-Leu-Phe-Gln↓Gly-Pro	high specificity, efficient at low temperatures
TagZyme (cysteine exopeptidase)	Stop Positions: ↓Lys; ↓Arg; ↓X-Pro; ↓X-X-Pro; ↓Gln [a]	sequential cleavage of dipeptides from the N-terminus, cleaves only small tags (up to 25aa), no non-specific cleavage

↓ – position of cleavage; X – any amino acid; [a] in the presence of excess Qcyclase; refer the text for details

Table 3. Comparison of commonly used expression systems (Yin et al., 2007)

Host cell	Cell growth time	Features	Post-translational modifications [a]					
			N-glyc.	O-glyc.	phosph.	acetyl.	acyl.	γ-carb.
E. coli	rapid (30 min)	easy operation and scale-up low cost and time, high yield problems with protein solubility	no	no	no	no	no	no
Yeast	rapid (90 min)	eukaryotic protein processing scalable up to fermentation simple media requirements	yes	yes	yes	yes	yes	no
Insect cells	slow (18-24 h)	near mammalian protein processing higher yield than mammalian system more demanding culture conditions	yes	yes	yes	yes	yes	no
Mammalian cells	slow (24 h)	mammalian protein processing relatively low yield and high cost demanding culture conditions	yes	yes	yes	yes	yes	yes

[a] N-linked glycosylation; O-linked glycosylation; phosphorylation; acetylation; acylation; gammacarboxylation

tion site or at least very close to its C terminus. Thanks to this feature it is possible to restore the exact sequence of the protein N terminus (with no additional amino acids). However, this is one reason why affinity tags are usually N-terminal, even though both ends are suitable for affinity purification purposes. After the cleavage of C terminal tags, the recognition sequences remain in the target protein.

Unfortunately, a number of problems are associated with the use of proteases: cleavage may be incomplete or proteins may be even resistant to it, and non-specific digestion may also occur. Both enterokinase and factor Xa, although they should generate proteins with native N-termini, often cleave at locations other than the desired site (Choi et al., 2001; Jenny et al., 2003). The latter phenomenon is often observed for enterokinase which recognizes the charge density rather than a particular amino acid sequence. Another disadvantage of factor Xa may be that this enzyme is not produced as a recombinant protein, but it is isolated from mammalian

plasma. This may create certain problems depending on the final use of the product. Also, thrombin exhibits non-specific cleavage as it does not recognize a long-defined specific sequence, but digests a variety of amino acids motifs. Some advantages of this enzyme in HTPP projects are its low cost and high efficiency of the cleavage. In the case of TEV and HRV specificity is much higher and cleavage occurs within the recognized sequence, not beyond it. However, TEV protease may tolerate a variety of residues in the position that remains at the protein N-terminus, allowing the production of native ends in many cases (Kapust et al., 2001; Kapust et al., 2002). HRV does not have such flexibility and leaves two strictly defined amino acids at the protein end (Cordingley et al., 1990).

An alternative approach to affinity tag removal is the use of exopeptidase. The most commonly used system: TagZyme is based on the activity of a recombinant dipeptide aminopeptidase I (DAPase) that sequentially cleaves dipeptides from the N terminus of virtually any protein (Arnau et al., 2008). The cleavage proceeds until a certain stop point in the sequence is reached. It may be N-terminal arginine, lysine or glutamine as well as proline at the second or third position ahead. DAPase may be used alone (if the target protein N-terminal sequence meets any of the conditions above) or in combination with two accessory enzymes. In the latter case, glutamine, that plays a role of a stop point, needs to be included between the tag and the protein sequence. TagZyme cleavage is followed by cyclization of the remaining N-terminal glutamine residue to pyroglutamate (catalyzed by Qcyclase, a glutamine cyclotransferase) and its subsequent removal (catalyzed by pGAPase, a pyroglutamyl aminopeptidase). In both the strategies, a tag-free protein with the native N terminus is obtained. This system may be adapted to a very efficient and precise removal of various short (up to 25 amino acids) N-terminal affinity tags. Its other advantage is the lack of non-specific digestion.

Although a large variety of affinity tags and their removal methods are available, the approach chosen by most HTPP laboratories is very similar in the outline. N-terminal hexahistidine tag has become a primary choice for expression and purification screening purposes. This choice is based on a general finding that there is no tag that can always guarantee highly effective protein production (Graslund et al., 2008). Although some tags have been shown to increase the overall solubility of fusion protein, their subsequent cleavage may results in target protein precipitation. In this context, His-tag is used to identify the targets that are easily expressed in a soluble form and may be subjected to straightforward and relatively cost-effective procedure of nickel affinity purification (so-called "low hanging fruit" proteins). MBP and GST are the second choice tags in the case of proteins previously identified as insoluble, whereas C-terminal His-tag is sometimes used when the N-terminal did not bind to an affinity column or could not be cleaved completely (Alzari et al., 2006; Kim et al., 2008). According to the Structural Genomics Consortium (SGC), the change in His-tag position from N to C terminus was helpful in 9% cases while fusions with large tags provided an additional 3% of successfully purified and crystallized proteins. One of advantages of hexahistidine tag is that sometimes there is no need to cleave it off as it rarely affects biological activity. Hexahistidine tag was even observed to support the crystallization process in some cases (Savitsky et al., 2010). When an affinity tag is to be removed, TEV protease seems to be the most suitable choice due to its high specificity, activity on a variety of substrates, and the efficient cleavage under a wide range of conditions (e.g. low temperature, broad range of pH, high ionic strength). His-tagged TEV protease may be easily produced in large quantities, reducing the overall costs of the tag removal process, which is essential for HTPP projects (Tropea et al., 2009).

Expression systems for protein production

A wide selection of expression systems for recombinant protein production is available, including bacteria, baculovirus-mediated insect cells, yeast, and several mammalian-based systems. There are significant differences among all of them in terms of proper folding and post-translational modifications of recombinant proteins as well as cost-effectiveness, speed and efficiency of production (Braun and LaBaer, 2003; Yin et al., 2007).

Escherichia coli is by far the most widely used host for recombinant protein synthesis. The organism is easy to manipulate, inexpensive to culture, and has a rapid growth rate. Moreover, it may produce recombinant protein extremely effectively (in amounts of up to 50% of the total bacterial proteins). However, as a prokaryotic system, it cannot perform the post-translational modifications that are sometimes relevant for the proper struc-

ture and function of eukaryotic proteins. Other most common problems that affect heterologously expressed proteins in bacteria are: incorrect folding (due to a lack of appropriate chaperon molecules), insolubility and formation of inclusion bodies (due to the high level of protein in the cell), and low efficiency of production (e.g. due to the differences in codon usage between the bacteria and the source organism of the target protein).

The two most important factors that influence total protein production are the type of promoter and the type of bacterial strain to be used. The most common promoters are: the T7 promoter which originates from bacteriophage T7 and the synthetic *trc* promoter derived from the promoters of *E. coli trp* and *lac* genes (Tegel et al., 2011). T7 promoter is the strongest one as it utilizes bacteriophage polymerase which is approximately five-fold more processive than *E. coli* RNA polymerase (Golomb and Chamberlin, 1974). To use this system, T7 polymerase gene needs to be present in the host strain chromosome (usually in the form of DE3 lysogen under IPTG-inducible *lac* UV5 promoter). The main difference between T7 and *trc* promoters is the level of basal expression of the target gene (up to 50% and 30% of the total cell protein, respectively) which is particularly important when the protein of interest is harmful to the host cell. While *trc* promoter exhibits high basal transcription, the T7 promoter is known for just a small leakage that can be further reduced using specially designed *E. coli* strains. Specific properties of the host strain are other important factors that may facilitate effective protein production (Samuelson, 2011). One of the most commonly used *E. coli* strains is BL21 – deficient in two bacterial proteases, which results in a reduced degradation of the recombinant protein. Various specific derivatives of this BL21 strain are available including strains that enhance cytoplasmic disulfide bond formation, compensate differences in codon usage, favor production of membrane proteins, stabilize genes containing repetitive sequences or enable much tighter control of the expression level. Special strains are also designed to label recombinant proteins with ^{35}S-methionine and selenomethionine for crystallography purposes. Other important parameters that may greatly influence the expression are medium formulation, temperature and inducer concentration. This large variety of possible combinations may be tested in parallel, small scale experiments resulting in the determination of optimum conditions for the production of each target protein (Peti and Page, 2007).

The eukaryotic expression system based on the yeast cells (for instance, *Pichia pastoris* or *Saccharomyces cerevisiae*) possesses many of the advantageous properties offered by *E. coli*: culture simplicity, rapid growth, and relatively low production costs. Additionally, as a eukaryotic system, it provides efficient heterologous protein folding, correct disulfide bonds forming, and some post-translational modifications. Unfortunately, the latter significantly differ from the mammalian modifications in the way N- and O-linked oligosaccharides are formed (Kukuruzinska et al., 1987; Kornfeld and Kornfeld, 1985). A very potent advantage is the possibility to equip the recombinant protein with a signal sequence that facilitates its secretion to the culture medium. This system was shown to be scalable up to a large fermentors format. Finally, yeast, being a food organism, is much easily acceptable for the production of pharmaceuticals as compared to *E. coli* (Idiris et al., 2010).

Another very powerful eukaryotic expression system is provided by baculovirus-infected insect cells. Its main advantage is the ability to carry out protein modification and processing in a way very similar (yet still not identical) to higher eukaryotes. However, specially designed insect cell lines were developed to produce humanized recombinant glycoproteins (Jarvis, 2003). Baculovirus is a vector of choice for heterologous gene delivery owing to its high capability, no infectivity to vertebrates and overall efficiency. This virus originally contains strongly expressed, late, and not-essential (in laboratory conditions) genes – polyhedrin and p10 genes. Therefore, hypertranscribed polyhedrin or p10 promoter is usually used for the synthesis of heterologous proteins in amounts of up to 30% of the total cell proteins. Late expressions may facilitate efficient production of toxic proteins; however, very late expressed proteins may not be fully modified. The reason for this may be that insect cell functions are declined during very late stage of infection, which eventually leads to cell death. Thus, contrary to prokaryotic and yeast systems, protein production, although efficient and scalable, cannot be run continuously. Finally, the expression in mammalian cells becomes an alternative when accurate post-translational modifications play a crucial role in proper folding and functioning of the target protein. Two basic strategies are used. The first one is transient gene expression and

the second one is stable gene expression (Wurm and Bernard, 1999). In the case of the former strategy, higher yields of protein may be obtained but the production is limited to a relatively short period of time. Human embryonic kidney (HEK) 293, baby hamster kidney (BHK), COS cells and their derivatives are commonly used as transient expression hosts owing to their high transfection efficiencies, ease of adaptation to serum-free suspension cultivation, and ability to cost-effective production scale-up. Other mammalian cell expression systems employ modified viruses as carriers of the target gene. The most popular are Semliki Forest Virus-, Vaccinia Virus-, and some Baculo- and Retrovirus-based vectors. The choice of a specific vector and cell line depends on the application, desired level of gene expression, type of post-translational modifications required and safety issues (Baldi et al., 2007; Van et al., 2000). The second strategy (stable gene expression), requires stable integration of foreign DNA into the host cell genome. Usually a lot of time and effort is required to select a proper cell line for stable gene expression. In such a system the level of target gene expression is often much lower than in transient expression systems, but it allows long-term protein production (Xia et al., 2006).

An interesting alternative for protein production in live cells is a cell-free system. It utilizes a crude cell extract enriched with all necessary components for *in vitro* protein synthesis based on RNA template. Such a system offers an ability to synthesize proteins that are toxic to the host cells. Moreover, the whole process may be easily controlled and automated, making it ideal for high throughput expression screening purposes. Among these – *E. coli*, rabbit reticulocyte – and wheat germ-based systems are the most widely used (Endo and Sawasaki, 2006; Vinarov and Markley, 2005).

High throughput screening may be applied to test various promoters, bacterial strains, cell lines and expression conditions, allowing the establishment of the best strategy for effective production of soluble proteins. It has been demonstrated that parallel processing of even a modest number of constructs can significantly improve the efficiency of gene expression (Savitsky et al., 2010). Nearly all structural genomics centers employ *E. coli* as their main expression host, reporting up to 70% of screening success (in terms of soluble protein production), largely dependent on protein size and origin (Braun and LaBaer, 2003; Alzari et al., 2006). Moreover, 90% of PDB deposited structures were determined on the basis of the proteins obtained in prokaryotic expression systems. While the major drawback of *E. coli* system is the lack of post-translational protein modification, its occurrence is particularly important for some membrane or secretory proteins. Thus eukaryotic expression is as of now reserved for high-value targets that are either not properly expressed in the prokaryotic system or are known to owe their functionality to specific modifications e.g. a specific glycosylation pattern. Currently, insect, yeast and mammalian cells are the key hosts used in a vast majority of eukaryotic expression systems (based on PDB data) in very general proportion of 3 : 2 : 1, respectively. Baculovirus-infected insect cells seem to be preferred as the eukaryotic system as they combine relatively good success rate, efficiency and cost-effectiveness, with post-translational modifications that are highly similar to mammalian cells. Transient expression in mammalian cells provides accurate post-translational modifications, correct folding and high success rates but is considered to be a system of the highest expense. However, recent studies that focus on structural genonics are changing this system into a robust and cost-effective alternative for efficient small scale expression screening (Aricescu et al., 2006a; Aricescu et al., 2006b; Banci et al., 2006).

Automation

A major feature of HTPP is the parallel processing of multiple probes. The whole production pipeline (as shown in Fig. 1) includes: amplification of cDNA, cDNA cloning, gene expression, analysis of protein solubility and protein purification trials. Each of these processes is typically adapted to a multiwell plate format and is combined with a certain automation method. While the scale of each step is reduced to not more than a few mililiters, the throughput is drastically increased. This saves time and is cost-effective, eliminates human errors, and guarantees reproducibility of the results.

For DNA procedures, 96-well plates are typically used, while for small scale expression screens plates of lower throughput (48, 24 wells) but higher volume are preferred (Aricescu et al., 2006b; Abdullah et al., 2009). Small scale purification has also been adapted to the high throughput format: affinity target (e.g. Ni ions) may be immobilized on magnetic beads compatible with automation facilities, or even directly on the wells surface.

A variety of liquid handling robots equipped with heater blocks, sample holders, vacuum manifolds, shakers or sonicators are available to operate on multiwell plates. Complete automation packages comprising a liquid-handling robot, reagents, and consumables are available from many manufacturers (e.g. Qiagen, GE Healthcare), which reduces the setup time but may raise the overall costs of processing. On the other hand, some laboratories have developed custom-built robotics for 96-sample bacterial fermentation and purification (Lesley, 2001; Chesneau et al., 2008).

The highest degree of automation combined with the largest range of screening variants may be achieved with *E. coli* expression system used for protein production. A straightforward automated procedure may involve bacterial cell culturing, expression induction, cell lysis, total protein isolation, and purification. The amount of soluble protein competent for the purification procedure may be easily estimated using UV absorption spectroscopy or polyacrylamide gel electrophoresis (Acton et al., 2005; Elsliger et al., 2010). Auto-inducting media provide an easy and efficient expression protocol since the cultures have only to be inoculated and grown to saturation. Moreover, the yields of target protein are usually higher than those obtained by conventional IPTG induction (Studier, 2005).

Prospects

The overall effort of numerous structural genomics laboratories is leading towards the determination of a complete set of protein folds. The aims are to provide more accurate information about the structures and functions of all known proteins as this knowledge is important for life sciences, biotechnology, and drug development. An efficient production of hundreds of proteins in parallel was greatly facilitated by the adoption of high throughput expression and purification techniques and thanks to the availability of genome sequencing data. Nowadays every single laboratory may be equipped in a way that permits rapid analysis of many constructs and many expression conditions, assuring efficient production of multiple proteins. Yet, some major challenges are still have to be solved. Some proteins, complexes and cellular assemblies are not compatible with the current high throughput pipelines. Membrane proteins, which represent about one-third of the proteins encoded by the human genome, appear to be the greatest challenge for high throughput technologies. Membrane proteins are very important targets for drug design, but their low natural abundance and general instability currently impedes their efficient production.

Acknowledgments

The publication of this article was supported by the project Structural Biology of Plants and Microbes, granted by the Foundation for Polish Science (project no. MPD/2008/2) and co-funded by European Union within European Regional Developmant Fund.

References

Abdullah J.M., Joachimiak A., Collart F.R. (2009) *"System 48" high-throughput cloning and protein expression analysis.* Meth. Mol. Biol. 498: 117-127.

Acton T.B., Gunsalus K.C., Xiao R., Ma L.C., Aramini J., Baran M.C., Chiang Y.W., Climent T., Cooper B., Denissova N.G. et al. (2005) *Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium.* Meth. Enzymol. 394: 210-243.

Alzari P.M., Berglund H., Berrow N.S., Blagova E., Busso D., Cambillau C., Campanacci V., Christodoulou E., Eiler S., Fogg M.J. et al. (2006) *Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE.* Acta Crystallogr. Sect. D: Biol. Crystallogr. 62: 1103-1113.

Aricescu A.R., Assenberg R., Bill R.M., Busso D., Chang V.T., Davis S.J., Dubrovsky A., Gustafsson L., Hedfalk K., Heinemann U. et al. (2006a) *Eukaryotic expression: developments for structural proteomics.* Acta Crystallogr. Sect. D: Biol. Crystallogr. 62: 1114-1124.

Aricescu A.R., Lu W., Jones E.Y. (2006b) *A time- and cost-efficient system for high-level protein production in mammalian cells.* Acta Crystallogr. Sect. D: Biol. Crystallogr. 62: 1243-1250.

Arnau J., Lauritzen C., Petersen G.E., Pedersen J. (2006) *Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins.* Protein Expr. Purif. 48: 1-13.

Arnau J., Lauritzen C., Petersen G.E., Pedersen J. (2008) *The use of TAGZyme for the efficient removal of N-terminal His-tags.* Meth. Mol. Biol. 421: 229-243.

Baldi L., Hacker D.L., Adam M., Wurm F.M. (2007) *Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives.* Biotechnol. Lett. 29: 677-684.

Banci L., Bertini I., Cusack S., de Jong R.N., Heinemann U., Jones E.Y., Kozielski F., Maskos K., Messerschmidt A., Owens R. et al. (2006) *First steps towards effective methods in exploiting high-throughput technologies for the determination of human protein structures of high biome-*

dical value. Acta Crystallogr. Sect. D: Biol. Crystallogr. 62: 1208-1217.

Benoit R.M., Wilhelm R.N., Scherer-Becker D., Ostermeier C. (2006) *An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids.* Protein Expr. Purif. 45: 66-71.

Bernard P. (1996) *Positive selection of recombinant DNA by CcdB*. Biotechniques 21: 320-323.

Berrow N.S., Alderton D., Owens R.J. (2009) *The precise engineering of expression vectors using high-throughput In-Fusion PCR cloning*. Meth. Mol. Biol. 498: 75-90.

Berrow N.S., Alderton D., Sainsbury S., Nettleship J., Assenberg R., Rahman N., Stuart D.I., Owens R.J. (2007) *A versatile ligation-independent cloning method suitable for high-throughput expression screening applications.* Nucl. Acids Res. 35: e45.

Blommel P.G., Martin P.A., Wrobel R.L., Steffen E., Fox B.G. (2006) *High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system*. Protein Expr. Purif. 47: 562-570.

Blommel P.G., Martin P.A., Seder K.D., Wrobel R.L., Fox B.G. (2009) *Flexi vector cloning.* Meth. Mol. Biol. 498: 55-73.

Braun P., LaBaer J. (2003) *High throughput protein production for functional proteomics*. Trends Biotechnol. 21: 383-388.

Chant A., Kraemer-Pecore C.M., Watkin R., Kneale G.G. (2005) *Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site*. Protein Expr. Purif. 39: 152-159.

Charlton A., Zachariou M. (2011) *Tag removal by site-specific cleavage of recombinant fusion proteins.* Meth. Mol. Biol. 681: 349-367.

Chesneau A., Yumerefendi H., Hart D.J. (2008) *The impact of protein expression methodologies on structural proteomics*. [in:] *Structural proteomics and its impact on the life sciences*, ed. Sussman J.L., Silman I., World Scientific Publishing Co. Pte. Ltd., Singapore: 207-232.

Choi S.I., Song H.W., Moon J.W., Seong B.L. (2001) *Recombinant enterokinase light chain with affinity tag: expression from Saccharomyces cerevisiae and its utilities in fusion protein technology*. Biotechnol. Bioeng. 75: 718-724.

Collins F.S., Morgan M., Patrinos A. (2003) *The Human Genome Project: lessons from large-scale biology.* Science 300: 286-290.

Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M., Colonno R.J. (1990) *Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro.* J. Biol. Chem. 265: 9062-9065.

Derewenda Z.S. (2004) *The use of recombinant methods and molecular engineering in protein crystallization*. Methods 34: 354-363.

Durbin R.M., Abecasis G.R., Altshuler D.L., Auton A., Brooks L.D., Durbin R.M., Gibbs R.A., Hurles M.E., McVean G.A. (2010) *A map of human genome variation from population-scale sequencing.* Nature 467: 1061-1073.

Elsliger M.A., Deacon A.M., Godzik A., Lesley S.A., Wooley J., Wuthrich K., Wilson I.A. (2010) *The JCSG high-throughput structural biology pipeline.* Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun. 66: 1137-1142.

Endo Y., Sawasaki T. (2006) *Cell-free expression systems for eukaryotic protein production*. Curr. Opin. Biotechnol. 17: 373-380.

Fonda I., Kenig M., Gaberc-Porekar V., Pristovaek P., Menart V. (2002) *Attachment of histidine tags to recombinant tumor necrosis factor-alpha drastically changes its properties*. Sci.World J. 2: 1312-1325.

Fong B.A., Wu W.Y., Wood D.W. (2010) *The potential role of self-cleaving purification tags in commercial-scale processes*. Trends Biotechnol. 28: 272-279.

Goel A., Colcher D., Koo J.S., Booth B.J., Pavlinkova G., Batra S.K. (2000) *Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct*. Biochim. Biophys. Acta 1523: 13-20.

Golomb M., Chamberlin M. (1974) *Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis*. J. Biol. Chem. 249: 2858-2863.

Gong W., Shen Y.P., Ma L.G., Pan Y., Du Y.L., Wang D.H., Yang J.Y., Hu L.D., Liu X.F., Dong C.X. et al. (2004) *Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes.* Plant Physiol. 135: 773-782.

Graslund S., Nordlund P., Weigelt J., Hallberg B.M., Bray J., Gileadi O., Knapp S., Oppermann U., Arrowsmith C., Hui R. et al. (2008) *Protein production and purification.* Nat. Meth. 5: 135-146.

Hammarstrom M., Hellgren N., van Den B.S., Berglund H., Hard T. (2002) *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli*. Protein Sci. 11: 313-321.

Hartley J.L., Temple G.F., Brasch M.A. (2000) *DNA cloning using in vitro site-specific recombination.* Genome Res. 10: 1788-1795.

Idiris A., Tohda H., Kumagai H., Takegawa K. (2010) *Engineering of protein secretion in yeast: strategies and impact on protein production*. Appl. Microbiol. Biotechnol. 86: 403-417.

Jarvis D.L. (2003) *Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production*. Virology 310: 1-7.

Jenny R.J., Mann K.G., Lundblad R.L. (2003) *A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa.* Protein Expr. Purif. 31: 1-11.

Kaplan W., Husler P., Klump H., Erhardt J., Sluis-Cremer N., Dirr H. (1997) *Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag*. Protein Sci. 6: 399-406.

Kapust R.B., Tozser J., Copeland T.D., Waugh D.S. (2002) *The P1′ specificity of tobacco etch virus protease.* Biochem. Biophys. Res. Commun. 294: 949-955.

Kapust R.B., Tozser J., Fox J.D., Anderson D.E., Cherry S., Copeland T.D., Waugh D.S. (2001) *Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency*. Protein Eng. 14: 993-1000.

Kim Y., Bigelow L., Borovilos M., Dementieva I., Duggan E., Eschenfeldt W., Hatzos C., Joachimiak G., Li H., Maltseva N. et al. (2008) *Chapter 3. High-throughput protein purification for x-ray crystallography and NMR.* Adv. Protein Chem. Struct. Biol. 75: 85-105.

Klock H.E., Lesley S.A. (2009) *The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis*. Meth. Mol. Biol. 498: 91-103.

Kornfeld R., Kornfeld S. (1985) *Assembly of asparagine-linked oligosaccharides.* Ann. Rev. Biochem. 54: 631-664.

Kou G., Shi S., Wang H., Tan M., Xue J., Zhang D., Hou S., Qian W., Wang S., Dai J. et al. (2007) *Preparation and characterization of recombinant protein ScFv(CD11c)-TRP2 for tumor therapy from inclusion bodies in Escherichia coli*. Protein Expr. Purif. 52: 131-138.

Kukuruzinska M.A., Bergh M.L., Jackson B.J. (1987) *Protein glycosylation in yeast.* Ann. Rev. Biochem. 56: 915-944.

Lamesch P., Li N., Milstein S., Fan C., Hao T., Szabo G., Hu Z., Venkatesan K., Bethel G., Martin P. et al. (2007) *hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes.* Genomics 89: 307-315.

Lamesch P., Milstein S., Hao T., Rosenberg J., Li N., Sequerra R., Bosak S., Doucette-Stamm L., Vandenhaute J., Hill D.E. et al. (2004) *C. elegans ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions.* Genome Res. 14: 2064-2069.

Lesley S.A. (2001) *High-throughput proteomics: protein expression and purification in the postgenomic world*. Protein Expr. Purif. 22: 159-164.

Liu Q., Li M.Z, Leibham D., Cortez D., Elledge S.J. (1998) *The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes*. Curr. Biol. 8: 1300-1309.

Luan C.H., Qiu S., Finley J.B., Carson M., Gray R.J., Huang W., Johnson D., Tsao J., Reboul J., Vaglio P. et al. (2004) *High-throughput expression of C. elegans proteins.* Genome Res. 14: 2102-2110.

Mayer A., Sharma S.K., Tolner B., Minton N.P., Purdy D., Amlot P., Tharakan G., Begent R.H., Chester K.A. (2004) *Modifying an immunogenic epitope on a therapeutic protein: a step towards an improved system for antibody-directed enzyme prodrug therapy (ADEPT).* Br. J. Cancer 90: 2402-2410.

Nagase T., Yamakawa H., Tadokoro S., Nakajima D., Inoue S., Yamaguchi K., Itokawa Y., Kikuno R.F., Koga H., Ohara O. (2008) *Exploration of human ORFeome: high-throughput preparation of ORF clones and efficient characterization of their protein products*. DNA Res. 15: 137-149.

Olsen D.B., Eckstein F. (1989) *Incomplete primer extension during in vitro DNA amplification catalyzed by Taq polymerase; exploitation for DNA sequencing*. Nucl. Acids Res. 17: 9613-9620.

Peti W., Page R. (2007) *Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost*. Protein Expr. Purif. 51: 1-10.

Rajan S.S., Lackland H., Stein S., Denhardt D.T. (1998) *Presence of an N-terminal polyhistidine tag facilitates stable expression of an otherwise unstable N-terminal domain of mouse tissue inhibitor of metalloproteinase-1 in Escherichia coli*. Protein Expr. Purif. 13: 67-72.

Samuelson J.C. (2011) *Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations*. Meth. Mol. Biol. 705: 195-209.

Savitsky P., Bray J., Cooper C.D., Marsden B.D., Mahajan P., Burgess-Brown N.A., Gileadi O. (2010) *High-throughput production of human proteins for crystallization: the SGC experience.* J. Struct. Biol. 172: 3-13.

Sievert V., Ergin A., Bussow K. (2008) *High throughput cloning with restriction enzymes.* Meth. Mol. Biol. 426: 163-173.

Smyth D.R., Mrozkiewicz M.K., McGrath W.J., Listwan P., Kobe B. (2003) *Crystal structures of fusion proteins with large-affinity tags*. Protein Sci. 12: 1313-1322.

Studier F.W. (2005) *Protein production by auto-induction in high density shaking cultures*. Protein Expr. Purif. 41: 207-234.

Tegel H., Ottosson J., Hober S. (2011) *Enhancing the protein production levels in Escherichia coli with a strong promoter*. FEBS J. 278: 729-739.

Tropea J.E., Cherry S., Waugh D.S. (2009) *Expression and purification of soluble His(6)-tagged TEV protease.* Meth. Mol. Biol. 498: 297-307.

Van C.K., Vanhoenacker P., Haegeman G. (2000) *Episomal vectors for gene expression in mammalian cells.* Eur. J. Biochem. 267: 5665-5678.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A. et al. (2001) *The sequence of the human genome.* Science 291: 1304-1351.

Vinarov D.A., Markley J.L. (2005) *High-throughput automated platform for nuclear magnetic resonance-based structural proteomics*. Expert Rev. Proteomics 2: 49-55.

Waugh D.S. (2005) *Making the most of affinity tags.* Trends Biotechnol. 23: 316-320.

Wurm F., Bernard A. (1999) *Large-scale transient expression in mammalian cells for recombinant protein production*. Curr. Opin. Biotechnol. 10: 156-159.

Xia W., Bringmann P., McClary J., Jones P.P., Manzana W., Zhu Y., Wang S., Liu Y., Harvey S., Madlansacay M.R. et al. (2006) *High levels of protein expression using different mammalian CMV promoters in several cell lines.* Protein Expr. Purif. 45: 115-124.

Yin J., Li G., Ren X., Herrler G. (2007) *Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes*. J. Biotechnol. 127: 335-347.