REVIEW
Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*

Per Jonasson*†, Sissela Liljeqvist*, Per-Åke Nygren* and Stefan Ståhl*[1]

**Department of Biotechnology, Kungliga Tekniska Högskolan, Stockholm, Sweden, and †AstraZeneca Biotech Laboratory, EST-Bio, R&D Södertälje, Sweden*

Genetic strategies have been used for more than two decades to improve bacterial bioprocesses and to simplify recovery procedures. Such strategies include the design of efficient expression vectors and the improvement of bacterial production strains in different ways, e.g. by deletion of protease genes or engineering for overexpression of rare-codon tRNAs, foldases or chaperones. Gene multimerization is another such principle that has proved beneficial to improve production yields. Genetic strategies have furthermore been exploited to facilitate recovery processes by adapting the product for a particular purification principle. In this area, affinity fusions have been commonly used, but other principles, such as modified isoelectric point (pI) or hydrophobic properties have also been successfully investigated. A recent drastic step forward in the use of gene technology to improve recovery processes for recombinant proteins is the introduction of combinatorial protein engineering to generate tailor-made product-specific affinity ligands. This strategy, which allows efficient recovery of a recombinant protein in its native form, is likely to be increasingly used also in industrial-scale bioprocesses, since novel protein ligands have been described that can be sanitized using common industrial cleaning-in-place procedures. The examples presented in this review make it evident that genetic strategies will be of utmost importance in the future for facilitating production and recovery of recombinant proteins.

Introduction

Recombinant DNA techniques used for obtaining and combinding genes from a variety of sources, and the possibility of expressing these genes in different host cells, have provided scientists in the biopharmaceutical field with proteins in quantities previously impossible to obtain. The choice of host for the production depends mainly on the properties and the final use of the expressed protein. If the protein consists of multiple subunits or requires substantial post-translational modifications, the preferred host usually is of higher eukaryotic origin. The bacterium *Escherichia coli* has, however, also been successfully used for production of relatively complex proteins [1], and progress over recent years has widened the use of this organism even further. As a production host, *E. coli* has mainly been used for cost-efficient production of large amounts of proteins that are limited in size and have a relatively simple structure. In addition to ensuring a large supply of the desired protein, the recombinant production of proteins also provides the opportunity to genetically design the product in order to facilitate the bioprocessing. Genetic-design approaches may be applied to influence the targeting of the gene product, i.e. whether the gene product is accumulated intracellularly in a soluble form or as inclusion bodies, or whether the product is secreted into the periplasm or even into the culture medium [2]. The selected production strategy can significantly influence the design of the purification scheme, as well as the quality and final yield of the protein produced. One example of how genetic design can be employed to facilitate the recovery of recombinant proteins is the use of gene fusions [3,4]. Creating fusion proteins may simplify the recovery process in such a way that it might be possible to integrate several unit operations, thereby increasing the overall efficiency in the downstream purification process. The hosts used for production of recombinant proteins range from simple prokaryotic organisms (bacteria) to multicellular organisms such as transgenic plants and animals, and including unicellular eukaryotic organisms such as yeast and the more complex eukaryotic insect and mammalian cells. The choice of host system depends on many

Key words: affibody, affinity tag, fusion protein, *in vitro* refolding, soluble expression.

Abbreviations used: IPTG, isopropyl β-D-thiogalactopyranoside; SPA, *Staphylococcus aureus* Protein A; RBS, ribosomal binding sites; ZZ, IgG-binding domains derived from staphylococcal protein A; BB, albumin-binding protein from streptococcal protein G; RSV, respiratory syncytial virus; EBA, expanded-bed adsorption; IMAC, immobilized-metal-ion-affinity chromatography.

[1] To whom correspondence should be sent, at the following address: Center for Molecular Biotechnology, Department of Biotechnology, SCFAB, KTH, SE-106 91 Stockholm, Sweden (e-mail stefans@biochem.kth.se).

factors, such as the size, structure and stability of the gene product, and the requirements for post-translational modifications for biological activity. The necessary production yields, acceptable cost and quality specifications of the final product also have to be considered.

E. coli as production host

When setting up a process for production of a recombinant protein, the normal approach is to first try to express the protein of interest in *E. coli*. Alternative expression systems are used only if the product is biologically inactive after production due to lack of essential post-translational modifications, incorrect folding or when the recovery of the native protein is too low [5]. Parameters important for successful production of a recombinant protein in *E. coli* include transcriptional and translational efficiency, stability of the expression vector and of the transcribed mRNA, localization, proteolytic stability and folding of the gene product, as well as cell growth [6,7]. The compartmental localization of the gene product and the related issues, such as proteolytic degradation and refolding, are discussed below.

The expression vector

An *E. coli* expression vector should contain, apart from the gene of interest, an origin of replication, a gene that confers antibiotic resistance (or an alternative selectable marker), a promoter and a transcription terminator. The origin of replication determines the vector copy number, which could typically be in the range of 25–50 copies/cell if the expression vector is derived from the low-copy-number plasmid pBR322, or between 150 and 200 copies/cell if derived from the high-copy-number plasmid pUC [8]. The copy number influences the plasmid stability, i.e. the maintenance of the plasmid within the cells during cell division. A positive effect of a high copy number is the greater stability of the plasmid when the random partitioning occurs at cell division. On the other hand, a high number of plasmids generally decreases the growth rate, thus possibly allowing for cells with few plasmids to dominate the culture, their being the faster growing [9]. Generally there appears to be no significant advantage of using higher-copy-number plasmids over pBR322-based vectors in terms of production yields [10].

The gene coding for antibiotic resistance is necessary both for identifying transformants and to ensure antibiotic selective pressure, that is, only cells that harbour an expression vector will divide, thus preventing plasmid loss. Genes conferring ampicillin, tetracyclin or kanamycin resistance are commonly used in expression vectors. Ampicillin resistance is mostly used only on a laboratory scale, because β-lactamase, the enzyme which confers the resistance, degrades ampicillin and thus the selective pressure is lost after a few generations of cell growth [5]. Furthermore, ampicillin has been thought to be potentially allergenic, and is therefore usually not the antibiotic of choice in the production of biotherapeutics intended for human use. Another approach in preventing plasmid loss is to use a mutated *E. coli* strain deficient in a gene encoding an essential protein, and include that crucial gene in the plasmid instead [7,9].

Table 1 A selection of promoters commonly used for high-level expression in *E. coli*

Promoter	Induction[a]	Reference(s)
lac	IPTG	[11]
tac	IPTG	[12,13]
trc	IPTG	[14]
T7	IPTG	[15,16]
trp	Trp starvation/β-IAA[b]	[10]
araBAD	L-Arabinose	[17]
$P_L(\lambda)$	Thermal	[18]
$P_R(\lambda)$	Thermal	[19]
lac(TS)	Thermal	[20]
P_{SPA}	Constitutive	[21]

[a]Most commonly used.
[b]Abbreviation: β-IAA, β-indoleacrylic acid.

A number of strong promoters are available for high-level expression in *E. coli* (Table 1). An important criterion of a promoter is its ability to be efficiently down-regulated under non-induced conditions, i.e. tightly regulated. An early overproduction of the heterologous protein, due to a non-silent promoter, might impair cell growth. It is therefore desirable to be able to repress the promoter during a cell growth phase to achieve high cell densities, after which the high-rate protein production would be initiated by induction of the promoter. Another important characteristic of a promoter is that it should be simple and inexpensive to induce. For laboratory-scale production, the isopropyl β-D-thiogalactopyranoside (IPTG)-inducible promoters, which are regulated by the product of the *lacI* gene, the lac repressor, are widely used. They include the *lac* promoter [11], the *lac–trp* hybrid promoter *tac* [12,13] and the *trc* promoter [14]. A disadvantage with these promoters is that they are not completely down-regulated under non-induced conditions, and thus are not suitable if the target-gene product is toxic to the cell. Another, more tightly regulated, IPTG-inducible expression system is the pET vector system, which is the most common expression system used in laboratory-scale cultivations [15,16]. The pET vector has a T7 promoter which is transcribed only by T7 RNA polymerase and must be used in a strain carrying a chromosomal T7 RNA polymerase gene which is under the control of a *lac* promoter. The use of IPTG for induction of these promoters might not still be optimal for the large-scale production of human therapeutic proteins because of the

cost of IPTG [22]. Lactose has been shown to be an inexpensive, but somewhat weaker, alternative for induction of the *lac* promoter in some applications [23]. For large-scale cultivations, either the *trp* promoter [10] or heat-induced promoters are commonly used. The *trp* promoter is induced by starvation of tryptophan or by the addition of β-indoleacrylic acid. One potential problem with the *trp* promoter is that it is difficult to completely down-regulate under non-induced conditions, a problem which, however, can be minimized by the addition of fructose to the cultivation medium [24]. Examples of heat-induced promoters are $P_L(\lambda)$ [18], $P_R(\lambda)$ [19] and the thermosensitive *lac* promoter *lac*(TS), which was constructed by mutation of the *lacI* gene [20]. One drawback with these promoters is that the thermal induction could also induce the production of heat-shock proteins, including certain proteases that can cause enhanced degradation. Constitutive promoters, such as the *Staphylococcus aureus* Protein A (SPA) promoter (P_{SPA}), have also been used for recombinant-protein production [21]. Promoters induced by cultivation conditions such as pH, oxygen levels, stationary growth and osmolarity [6], as well as weak and moderately strong promoters [7,25], have also been described.

A transcription termination downstream of the coding sequence has been described as enhancing plasmid stability by preventing transcription through the replication region and through other promoters located on the plasmid [22]. In addition, the transcription terminator enhances the stability of the mRNA transcript by a stem–loop formation at the 3′ end [26,27]. The tandem T1T2 transcription terminator [28], derived from the *rrn*B ribosomal RNA operon of *E. coli*, is an efficient and commonly used transcription terminator [6].

Protein production

Translation is initiated by the binding of the ribosomes at the Shine–Dalgarno (SD) sequences located within the ribosomal binding sites (RBS) in the mRNA sequence. Optimal translation initiation is obtained for mRNAs with the Shine–Dalgarno sequence UAAGGAGG. Also the space between the binding site and the initiation codon, ideally four to eight nucleotides in length, is important for efficient translation initiation [29]. Furthermore, the secondary structure around the RBS and in the sequence immediately downstream of the start codon have been described to influence the translational initiation efficiency, and an enrichment of A and T residues in those regions has been shown to improve the efficiency of translation [30,31]. It has recently been suggested that the codon that follows the AUG initiation triplet (the +2 codon) is of particular importance for the translation initiation efficiency, and that there is a preference for adenine residues in this codon in highly expressed gene products [32].

Table 2 Rare codons in *E. coli*, which should be avoided in clusters

Codon	Encoded amino acid	Frequency in *E. coli* genes (%)
AGG	Arg	0.14
AGA	Arg	0.21
CGA	Arg	0.31
CUA	Leu	0.32
AUA	Ile	0.41
CCC	Pro	0.43

The frequencies with which the different codons appear in genes in *E. coli* are different from those in genes of human origin. The amount of specific tRNAs is also reflected by the frequency of the codon, which means that a tRNA which recognizes a rarely used codon is present in low amounts. Therefore, human genes that contain codons which are rare in *E. coli* may be inefficiently expressed. This problem can be solved either by exchanging codons in the target gene for codons which are more frequently used in *E. coli*, or, alternatively, by co-production of the rare tRNAs. The most abundant codons in *E. coli* have been determined by examination of sets of genes, and lists of codon usage can be found in several publications, for example [33]. The effect on expression levels by substitution of rare codons with optimal ones has been extensively studied, but general conclusions have been difficult to draw. In some studies, an increased production level has been reported, but according to the reports of others, no effect could be detected [6]. It has, however, been found that a long stretch of similar codons decreases the expression level [34]. Furthermore, clusters of very rare codons (Table 2) can create translation errors and reduce the expression level, suggesting these codons should be substituted [35–37].

The preferred stop codon in *E. coli* is UAA [38]. However, several consecutive stop codons have been included to obtain efficient translation termination [22]. Alternatively, the prolonged UAAU stop codon can be used for more efficient translational termination [39].

Cultivation

Batch cultivation is the most simple way to produce a recombinant protein. In a batch cultivation, all the nutrients required for cell growth are supplied from the start, and the growth is initially unrestricted. However, the unrestricted growth commonly leads to unfavourable changes in the growth medium, such as oxygen limitation and pH changes. Also, certain metabolic pathways in the cell will be saturated, which potentially leads to the accumulation of inhibitory by-products in the medium. Therefore only moderate cell densities and production levels can normally be obtained with batch cultivations [40]. To obtain high cell density and high protein production levels, fed-batch cultivation in a bioreactor is commonly used. In a fed-batch cultivation, the carbon/energy source is added in proportion

to the consumption rate. Thereby overflow metabolism and the accumulation of inhibitory by-products can be minimized [41]. Moreover, the growth rate can be balanced to achieve a maximal production level. The bioreactor should preferentially be equipped to maintain an optimal oxygen concentration, pH and temperature [42]. Defined media are generally used in fed-batch cultivation. As the concentrations of the nutrients are known and can be controlled during the cultivation, the cultivation is also more reproducible compared with the use of a complex growth medium. However, the addition of complex media, such as yeast extract, is sometimes necessary to obtain a high level of the desired recombinant protein.

Acetate is produced when the culture is growing in the presence of excess glucose or under oxygen-limiting conditions [43]. A high concentration of acetate reduces growth rate, maximum obtainable cell density and the level of production of the recombinant protein. It is therefore important to maintain the acetate concentration below the inhibitory level. This can be achieved by controlling the cultivation in several ways: the growth rate could be controlled by limiting nutrients, such as sources of carbon or nitrogen [44,45], by using glycerol [45] or fructose [46] instead of glucose as the carbon/energy source, by addition of glycine and methionine [47] or by lowering the cultivation temperature, or by metabolic engineering [48,49]. Other problems concerning growth to high cell densities are oxygen limitation, reduced mixing efficiency, heat generation and high partial pressure of CO_2 [40,44].

Strategies for production

A major consideration when designing a process for production of a recombinant protein in *E. coli* is whether the gene product should be produced intracellularly or if a secretion system could be used. Different genetic design strategies, together with the inherent properties of the target protein, decide which expression route will be the most successful. Upon intracellular expression, the product can either accumulate as a soluble gene product or precipitate in the form of inclusion bodies. If a secretion system is used, and the product is found to be secretable, the gene product will be accumulated in the periplasm or in some cases even be translocated also through the outer membrane to the extracellular culture medium [2,50]. Every production strategy has its advantages and disadvantages (Table 3).

Production by secretion

The periplasm contains only about 100 proteins as compared with about 4000 proteins in the cytoplasm [51]. Thus considerable purification and concentration effects are achieved by the targeting of the gene product to the periplasm. Additional beneficial effects achieved through the secretion of the gene product include enhanced disulphide-bond formation, possibility to obtain gene products with authentic N-termini, decreased proteolysis and minimization of harmful action of recombinant proteins which are deleterious to the cell. The specific release of the periplasmic protein content is simple and commonly used at the laboratory scale by different osmotic-shock procedures. However, on an industrial scale, efficient methods for selective release of periplasmic proteins are lacking. Nevertheless, it has been shown that treatment at an elevated temperature after completed cultivation can improve unspecific leakage to the culture medium [52]. It would be even more attractive to obtain translocation of the gene product to the growth medium, since this would lead to a significantly simplified purification scheme for the gene product. Protection against proteolysis might also be achieved using this strategy, because *E. coli* has very low extracellular proteolytic activity under normal conditions [5].

Secretion into the culture medium There are no efficient pathways available for specific translocation of proteins through the outer membrane of *E. coli*. Instead, the secretion of some recombinant proteins to the periplasm is suggested to cause a destabilization of the outer membrane, which becomes leaky and allows the protein to diffuse into the extracellular medium in a semi-specific manner. This highly protein-specific phenomenon is not fully understood [53,54]. Examples of proteins which have been efficiently secreted to the culture medium include different heterologous proteins fused to SPA domains [55], to calmodulin [56] and to the OmpA signal sequence [57]. Another strategy is to use leaky *E. coli* mutants which constitutively release periplasmic proteins into the culture medium due to loss of outer-membrane integrity [58]. However, these mutants are fragile and revert readily to the non-leaky phenotype, a fact which makes these strains unsuitable for large-scale protein production. Alternatives to the use of leaky mutants are co-expression of the bacteriocin release protein [59] or of the third topological domain of the transmembrane protein TolA [60], whereby a leaky phenotype is induced by disrupting the integrity of the outer membrane causing periplasmic proteins to leak into the growth medium. Supplementation of the growth medium with glycine has also been shown to enhance the release of periplasmic proteins into the cultivation medium [61,62].

Secretion to the periplasmic space Many recombinant proteins have been successfully secreted to the periplasm by fusion of a signal sequence or a normally secreted protein N-terminally to the target protein. Frequently used signal

Table 3 Advantages and disadvantages of different strategies for the production of recombinant proteins in *E. coli*

Production strategy	Advantages	Disadvantages
Secretion/leakage to the extracellular medium	Disulphide formation possible Extensive proteolysis might be avoided Possible to obtain authentic N-terminus Significantly reduced levels of contaminants No need for cell disruption	Secretion to the medium usually not possible Dilution of the product
Periplasmic production	Disulphide formation possible Possible to obtain authentic N-terminus Reduced levels of contaminants	Secretion to the periplasm not always possible No large-scale procedure for selective release of periplasmic proteins available Periplasmic proteases can cause proteolysis
Intracellular production as inclusion bodies	Inclusion bodies easy to isolate Protection from proteases Protein is inactive and cannot harm host High production yields usually obtained	Solubilizing and *in vitro* folding necessary which usually give lower yields and higher cost Normally no authentic N-terminus
Intracellular and soluble production	No need for solubilization and refolding	High level of intracellular product can be harmful to the cells Complex purification Proteolysis might occur Disulphide formation usually not possible Normally no authentic N-terminus

sequences include those derived from the *E. coli* periplasmic proteins PhoA and MalE, the outer membrane proteins OmpA and LamB [63], β-lactamase [64] and DsbA [65]. Interestingly, the Gram-positive signal sequence derived from SPA has shown to efficiently direct recombinant proteins to the periplasm of *E. coli* [19].

Proteolysis caused by envelope proteases is one of the most severe problems encountered when directing a recombinant protein to the periplasm of *E. coli*. The proteolysis can be minimized by different approaches, for example by using protease-deficient strains or by genetic design of the gene product [50,66]. Proteases which degrade many heterologous proteins in the periplasm are DegP, Tsp (denoted Prc in some publications), protease III (also named Pi) and OmpT [67,68]. *E. coli* strains with single, double and triple mutants of these proteases have been shown to efficiently decrease the degradation of different heterologous proteins secreted to the periplasm [66,69]. Furthermore, mutations in the *rpoH* gene, which code for the heat-shock-response σ^{32} factor, have been shown to increase the production of secreted proteins through decreased degradation [66]. One problem in using hosts deficient in multiple proteases is that viability, and thus growth, is impaired. Growth condition parameters such as temperature, pH and medium composition also affect the periplasmic proteolysis [69,70]. Genetic design approaches include *in vitro* mutagenesis in order to specifically eliminate protease cleavage sites in the target protein gene, and different fusion protein strategies to protect the target protein from proteolysis [50]. For example, the two IgG-binding domains ZZ, derived from SPA [71], and the albumin-binding protein BB from streptococcal protein G [72], have successfully been used as fusion partners serving this purpose. They have either been fused to the N-terminus, the C-terminus or to both termini of the target protein. The most pronounced stabilization effect has been obtained using the dual-affinity fusion strategy [73,74], which, in addition, allows recovery of the full-length product by two subsequent affinity-purification steps. The same effect has also been demonstrated using combinations of other affinity tags [3].

The environment in the periplasm is less reducing than that of the cytoplasm and favours the correct folding of recombinant proteins containing disulphide bonds. The periplasmic space also harbours foldases involved in the formation of disulphide bonds and isomerization of the proline imide bonds [75,76]. During recent years, several studies have been aimed at investigating whether co-expression of some of these foldases, or the eukaryotic equivalents thereof, enhances the yield of correctly folded protein in the *E. coli* periplasm. In some studies enhanced yields were achieved, but in others, only minor or no effects were observed [77–82]. Co-expression of a selective binder to the target protein is another approach to obtain enhanced yield of correctly folded protein. By this approach, co-expression of insulin-like growth factor-I with its binding protein, together with the addition of a redox buffer to the culture, resulted in almost quantitative yields of correctly folded insulin-like growth factor-I [83].

Intracellular production

An intracellularly produced recombinant protein can be accumulated in a soluble form in the cytoplasm, precipitate and form inclusion bodies, or, alternatively, be partly in the form of inclusion bodies and partly in soluble form. It is usually impossible to predict whether the gene product will be soluble or if it will precipitate [84], and empirical investigations are therefore necessary. Among the most important factors influencing the inclusion-body formation are protein expression rate and presence of disulphide

bonds, but hydrophobicity and choice of fusion partner have also been shown to have a significant impact [27].

Production of soluble gene products If the gene product is stable against proteolysis and not harmful to the host cell, it might be desirable to keep the protein soluble in the cytoplasm and thereby avoiding the solubilization and refolding steps that have to be performed if inclusion bodies are formed. There are several different approaches to minimize the formation of inclusion bodies when producing heterologous proteins intracellularly in *E. coli*. Reduction of the rate of protein synthesis, which can be achieved by using a moderately strong or weak promoter, or partial induction of a strong promoter, has been found to result in a higher amount of soluble protein [25]. Other means of reducing the protein-synthesis rate is by growing the culture at lower temperature [85] or to add non-metabolizable carbon sources at the time of induction [84]. Substitution of amino acid residues, for example replacement of multiple hydrophobic phenylalanine residues in respiratory-syncytial-virus (RSV) G protein [86], or replacement of cysteine residues in S1 dihydrofolate reductase [87], have been shown to dramatically improve the solubility. However, this approach is limited to applications where the substitutions do not alter the desired function or activity of the recombinant protein. Fusion of the target protein to a highly soluble fusion partner, thereby increasing the overall solubility of the fusion protein, is a convenient and efficient method to increase the fraction of soluble gene product in the cytoplasm. Proteins used as solubilizing fusion partners include thioredoxin [88,89], ubiquitin [90,91], NusA [92] the IgG-binding domains ZZ from SPA [93], the albumin-binding BB from Protein G [94–96], the maltose-binding protein [97,98] and a mutant form of DsbA [99]. The overexpression of intracellular chaperones has in many studies resulted in an increased accumulation of soluble gene products. However, as for co-expression of foldases, this approach is protein-specific and is not a universal means of preventing inclusion-body formation [75,77,78,80,81,100–102]. Usually the redox potential in the cytoplasm prevents disulphide formation. In order to generate a less reducing environment in the cytoplasm, thereby facilitating disulphide-bond formation, strains deficient in thioredoxin reductase have been used [103–105].

For soluble gene products accumulated intracellularly in the *E. coli* cytoplasm, the first step in downstream processing is the release of the recombinant protein. On a laboratory scale the cells are typically lysed by enzymic treatment, chemical treatment or by mechanical-disruption techniques such as sonication. High-pressure homogenization or bead mills are used in large-scale processing [106,107]. Such treatment effectively liberates the desired protein, but it also releases the bulk of host-cell proteins and nucleic acids. If the expressed recombinant protein is thermostable, a convenient method to reduce the amounts of the contaminating host-cell proteins is heat-precipitation [108,109]. Additional advantage with heat precipitation is the thermal deactivation of the *E. coli* cell and of its proteases [110,111], reducing the potential risk of degradation of the target protein. It has also been shown that heat-treatment procedures performed on undisrupted cells efficiently can release recombinant proteins accumulated in a soluble form in the cytoplasm [112], thus combining the product-release step with the benefits of heat precipitation of host-cell proteins.

Such a heat-treatment procedure was recently successfully used in the recovery of an intracellularly accumulated fusion protein, BB-C7, in a production process for human proinsulin C-peptide [96], where the heat treatment actually functioned as an initial purification step, giving a purity of approx. 70%, as compared with a purity of 10% obtained after conventional cell-disruption procedures.

Production as inclusion bodies Many heterologous proteins expressed in *E. coli* are prone to precipitate, which in many cases is an advantage. The formation of inclusion bodies normally protects the gene product from host-cell proteases. The product is inactive and cannot harm the host cell, often giving high expression levels. Furthermore, the dense inclusion bodies can be readily recovered by centrifugation and a relatively high purity and degree of concentration of the gene product are thus normally obtained after solubilization [1]. The main disadvantage with inclusion-body formation is the need for solubilization and refolding steps, necessary for regaining a correct protein structure and activity. These steps can reduce the yield and be costly, especially on a large scale. Different strategies have been utilized to enhance the tendency for the formation of inclusion bodies, for example increasing the rate of protein synthesis by using strong promoters such as the T7, *trp* or *tac* promoters, fusion of the target protein to certain other proteins, such as TrpLE [5,113], and cultivation at elevated temperatures or at a pH other than 7.0 [114,115].

In vitro refolding Inclusion bodies have an increased density and can easily be recovered by centrifugation after the disruption of the cells. The resulting inclusion-body-containing pellet consists mainly of the overexpressed recombinant protein, but contaminants originating from the host cells are also present [1]. To remove these contaminants, the pellet can be washed with low concentration of denaturants or with detergents [116]. After washing, the inclusion bodies are solubilized by using high concentration of denaturants. If the recombinant protein contains cysteine

residues, a reducing and chelating agent should also be included in the solubilization buffer [1,117].

Renaturation of the solubilized gene product is initiated by the removal of the denaturant and, where appropriate, also the reducing agent, by dialysis or dilution. During the refolding procedure it is important to limit product aggregation. This can be done by performing the refolding at low protein concentration, typically in the range of 10–50 mg/l [118]. However, refolding at such low concentrations requires very large volumes, which becomes difficult and expensive when performed in industrial-scale applications. Therefore other methods to keep the concentration of the unfolded protein low in the refolding buffer have been developed. Stepwise addition of the denatured recombinant protein and different dialysis approaches [116–118] are examples of such methods. Different strategies have been developed to increase the refolding yield, either by stabilizing the native state, by destabilizing incorrectly folded molecules, or by increasing the solubility of folding intermediates and of the unfolded state. By performing the refolding at non-denaturing concentrations of denaturant, a high refolding yield has been obtained at high protein concentrations [84,119]. Other low-molecular-mass additives have also successfully been used to enhance the refolding yield of a variety of different recombinant proteins [116]. Molecular chaperones and foldases [80,120,121], monoclonal antibodies [122,123] and specific binding proteins [124] have also been shown to increase the yield of correctly folded protein.

If the recombinant protein contains disulphide bonds, the renaturation buffer also has to contain a redox system, which provides the appropriate redox potential and enables formation and reshuffling of disulphides. The most common redox system is that of GSH and GSSG, but other low-molecular-mass thiol-based redox systems have also been utilized. Typically a 1:1–5:1 molar ratio of reduced to oxidized thiol is used [84]. For certain proteins, the yield of renaturation is increased if the thiol groups in the denatured protein are first completely oxidized by formation of mixed disulphides with GSH. Disulphide-bond formation is promoted by addition of catalytic amounts of a reducing agent in a following renaturation step [117]. In another method, the thiol groups in the denatured protein are sulphonated by treatment with Na_2SO_3 and a reducing agent. Under renaturating conditions the protein is thereafter refolded in the presence of small amounts of reducing agent [125–127].

Purification of the gene product

After a successful production of a recombinant protein, different purification steps will be needed in order to recover a biologically active protein at high purity. The downstream process for the recovery and purification of a gene product depends on the production strategy used, but it consists typically of product release and clarification steps, an initial purification step and different chromatographic purification steps. During recent years, the major challenge in designing downstream processes has been to simplify and improve the overall efficiency by combination and elimination of unit operations to cut production costs. This has been achieved both by the development of new separation techniques and by genetic design of the produced recombinant protein.

Initial recovery methods

The aim of an initial recovery step is to rapidly remove or inactivate proteases which can degrade the product, to remove impurities and particles which have a negative effect on the subsequent chromatographic purification steps and to concentrate the sample. An ideal initial recovery step also gives a high degree of purification. Furthermore, it is essential that the equipment used is compatible with robust cleaning and sanitizing methods when considering industrial-scale production.

The expanded-bed adsorption (EBA) technology represents an initial recovery step that allows the capture of proteins from particle-containing feedstocks without prior removal of the particulates [128]. EBA has shown to be suitable for industrial production scale and the technology can also withstand harsh cleaning procedures [129]. Precipitation is another simple approach for recovery of a gene product from a cultivation broth or homogenate. Various methods exist by which precipitation can be achieved: addition of salts, organic solvents, or organic polymers, or varying the pH or temperature [130]. These precipitation methods are non-specific, and give a low degree of purification. By using affinity precipitation, increased specificity can be obtained [131]. Different aqueous two-phase extraction systems have been extensively studied as an initial recovery step [132,133]. An aqueous two-phase extraction system can also be combined with affinity precipitation [134], combining the benefits of both methods.

Chromatographic purification methods

For recombinant proteins intended for use as pharmaceuticals, purity must typically exceed 99%, and some impurities, such as endotoxins and DNA, are limited to an upper level in the range of parts per million. Different chromatographic methods have proved to be the only purification techniques with which these purity levels can be obtained with retained biological activity of the product, suitable to be scaled up to the appropriate production levels. The most frequently used chromatographic methods are

Table 4 Commonly used affinity-fusion systems

Abbreviations: aa, amino acids; ABP, albumin-binding protein; GST, glutathione S-transferase; hIgG, human IgG; HSA, human serum albumin; mAb, monoclonal antibody; MBP, maltose-binding protein; Me^{2+}, bivalent metal ion; FLAG, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys.

Fusion partner	Size	Ligand	Elution[a]	Reference
Protein A	31 kDa	hIgG	Low pH	[19]
Z	7 kDa	hIgG	Low pH	[71]
ABP	5–25 kDa	HSA	Low pH	[72]
Hexahistidine	6 aa	Me^{2+} chelator	Imidazole/low pH	[139]
GST	25 kDa	GSH	GSH	[140]
MBP	41 kDa	Amylose	Maltose	[97]
FLAG peptide	8 aa	mAb M1	EDTA/low pH	[141]
		mAb M2	Low pH	[142]
Pin Point[b]	13 kDa	Streptavidin/avidin	Biotin	[143]
Bio[c]	13 aa	Streptavidin/avidin	Diaminobiotin	[144]

[a]Most common elution method.

[b]Subunit of the transcarboxylase complex from *Propionibacterium shermanii*, biotinylated *in vivo* by *E. coli*.

[c]Peptide selected from a combinatorial library and found to be biotinylated *in vivo*.

ion-exchange chromatography, size-exclusion chromatography, hydrophobic-interaction chromatography, reversed-phase chromatography and affinity chromatography.

In a typical multistep chromatography process, the first step is a capturing step, where the product binds to the adsorbent while the impurities do not. The product is often eluted with a step gradient, giving a high concentration of the product but a moderate degree of purification. The main requirements of this first step are high capacity, high degree of product recovery, and high chemical and physical stability. Ion-exchange chromatography, and to some extent hydrophobic-interaction chromatography, are frequently used as the first chromatographic step. After the capturing step, the bulk of impurities, such as host-cell proteins, nucleic acids and endotoxins, are typically removed with high-resolution techniques such as hydrophobic-interaction chromatography, ion-exchange chromatography, reversed-phase chromatography or affinity chromatography. Lower flow rates, gradient elution, and matrices with particles of smaller size are used for enhanced resolution. After these steps, the product purity is typically at a level of 99%. The last step in the multistep chromatography process is a polishing step, with the purpose of removing possible aggregates, degradation products or target protein molecules that may have been modified during the purification procedure. It also serves to condition the purified product for its use or storage. Commonly used techniques for the final step are size-exclusion chromatography and reversed-phase chromatography. When designing a purification scheme it is advisable to link the different chromatography steps in a way that enables the eluted sample from one step to be applied directly on to the next step, avoiding buffer changes and concentration steps. It is also important to keep the number of steps as low as possible, since the total recovery decreases rapidly with the increasing number of steps. A convenient way to reduce the number of steps in the purification scheme without reducing the purity is to include a step with high selectivity, such as affinity chromatography [135–137].

Affinity-fusion systems

Since the first example of the use of gene fusions for affinity purification was reported in 1983 [138], a large number of different affinity-fusion systems have been developed to facilitate the purification of recombinant proteins [3,4]. Some of the most common affinity-fusion systems are listed in Table 4. When choosing an affinity-fusion system, it is important to remember that all systems have their own characteristics, and no single system is ideal for all applications. For example, if secretion of the gene product is desired, it is necessary to choose a system with a secretable affinity tag. If the gene product needs to be purified under denaturated conditions, a system which has a tag which can bind under those conditions must be chosen, e.g. the polyhistidine affinity tag [139,145]. The polyhistidine tag is suitable for purification of gene products accumulated as inclusion bodies, because the fusion protein can be directly applied to an immobilized-metal-ion-affinity-chromatography (IMAC) column after being solubilized with a suitable denaturing agent. An additional advantage with the small polyhistidine affinity tag is that it can easily be genetically fused to a target protein by PCR techniques [146]. It is also important to choose an affinity-fusion system with elution conditions under which the target protein does not get denaturated [147]. At the laboratory scale, affinity-fusion methods are very powerful and have achieved widespread use for easy and fast single-step purification of gene products. For large-scale pharmaceutical production, however, affinity fusions have not been as extensively utilized, despite the ability to replace multiple steps with one step. The main reason is most probably that, for most applications, the affinity tag needs to be removed afterwards. Furthermore, proteinaceous ligands may leak from the column during elution, making it necessary to remove the ligand from the eluate. If the ligand originates from a mammalian source, there is also risk of viral contamination. Questions concerning the possibility of column sanitation, and column lifetime, capacity and cost, must also be considered [148].

Cleavage of fusion proteins For certain applications, for example if the produced protein is intended to be used as a pharmaceutical product, or if the protein is aimed to be used for structural determination, it is necessary to remove the affinity tag after the affinity-purification step. There are several methods, based on chemical or enzymic treatment, available for site-specific cleavage of fusion proteins, [149]

Table 5 Examples of methods used for site-specific cleavage of fusion proteins

Abbreviations: Xaa, unspecified amino acid; GST, glutathione S-transferase; ABP, albumin-binding protein; TEV, tobacco-etch virus; ↓, cleavage site; [H64A], $His^{64} \rightarrow Ala$; His_6, hexahistidine.

Cleavage agent	Cleavage sequence[a]	Reference
Chemical agents		
Hydroxylamine	Asn-↓-Gly	[150]
CNBr	Met-↓-Xaa	[151]
Enzymes		
Enterokinase	Asp-Asp-Asp-Asp-Lys-↓-Xaa	[141]
[H64A]Subtilisin	Phe-Ala-His-Tyr-↓-Xaa	[152]
IgA protease	Pro-Ala-Pro-Arg-Pro-Pro-↓-Thr	[153]
Factor Xa	Ile-Glu-Gly-Arg-↓-Xaa	[154]
Thrombin	Leu-Val-Pro-Arg-↓-Gly-Ser	[155]
GST–protease 3C	Leu-Glu-Val-Leu-Phe-Gln-↓-Gly- Pro	[156]
ABP–protease 3C–His_6	Leu-Glu-Ala-Leu-Phe-Gln-↓-Gly-Pro	[157]
His_6–TEV protease	Glu-Asn-Leu-Tyr-Phe-Gln-↓-Gly	[158]

[a]The cleavage sequence relates to the specific reference given; for some of the enzymes, alternative sequences are also reported to be functional.

(Table 5). Advantages with the chemical cleavage methods are that the reagents used are inexpensive and widely available, and the reactions are generally easy to scale up. However, the harsh reaction conditions often required can lead to amino-acid-side-chain modifications or denaturation of the target protein [159]. Furthermore, the selectivity is often rather poor, and cleavage can occur on additional sites within the target protein. Therefore chemical cleavage methods are usually only suitable for release of peptides and smaller proteins. For many applications, enzymic cleavage methods are preferred to chemical ones, because of their higher selectivity, and because the cleavage often can be performed under physiological conditions. Disadvantages of enzymic cleavage methods are that some enzymes are very expensive, and that not all enzymes are widely available. Furthermore, if the enzyme is of mammalian origin, virus-removal and virus-clearance validations need to be performed if the target protein is to be used as a pharmaceutical. Recombinant proteases, produced in bacteria or yeast, are for that reason preferred.

Genetic strategies to improve downstream processes

In recent years the number of reports describing the use of genetic strategies have increased, not only to improve production, but also to simplify the recovery processes. Such strategies have been used both to simplify initial recovery steps and to adapt the product to make the different purification steps more efficient. Genetic strategies have furthermore been used to facilitate *in vitro* refolding, site-specific cleavage of gene fusion product and to create tailor-made product-specific affinity ligands. Some relevant examples will be described below.

Improved recovery

Examples of gene fusions that have been used to improve initial recovery steps include fusions of hydrophobic tails to the target proteins to favour the partitioning into the top phase in aqueous two-phase systems [160,161], and fusions of aspartic acid residues to the protein to enhance polyelectrolyte precipitation efficiency [162]. Increased efficiency in anion-exchange chromatography in the EBA format was achieved by fusion of the target protein to the ZZ domains from Protein A, whereby the pI was lowered [52]. By fusion of a stretch of arginine residues [163], glutamic acid residues [164] and phenylalanine residues [165], the efficiency of ion-exchange chromatography and hydrophobic-interaction chromatography was increased. One example of a tailor-made fusion partner, developed by Gräslund and co-workers [166], is the engineered basic variant of the Z domain (Z_{basic}), enabling cation-exchange-chromatography separations to be performed at high pH values. Since almost no other host-cell proteins were found to bind under such conditions, very efficient purification could be achieved [167]. Utilizing the features of the charged Z_{basic}, an integrated production strategy for Klenow DNA polymerase was developed [168]. The Klenow DNA polymerase was produced as a Z_{basic}–Klenow fusion protein that could be efficiently recovered by cation-exchange chromatography in the EBA mode. The Z_{basic}–Klenow fusion was subsequently cleaved to release free Klenow polymerase, with the help of a Z_{basic}-tagged viral protease 3C, whereafter fused Klenow could be recovered from the reaction mix by separating Z_{basic}-protease 3C and Z_{basic} fusion partner using cation-exchange chromatography [168].

Facilitated *in vitro* refolding

Fusions of target proteins to highly soluble fusion tags have been shown to enhance *in vitro* refolding. For example, a high

refolding yield at high protein concentration was obtained by fusion of a moderately soluble target protein to ZZ from protein A [127,169,170]. By fusion of a target protein to a histidine tag, immobilization of the fusion protein on an IMAC column can be made under denaturating conditions. A subsequent on-column refolding step typically gives high yield of renaturated target protein [171]. A related example, in which a hexa-arginine polypeptide extension was fused to the target protein, the fusion protein was immobilized on a cation-exchange column and renatured target protein was obtained after on-column refolding [172].

Increased yield by gene multimerization

When expressing peptides in *E. coli*, low yields are often obtained. One reason could be the susceptibility of the peptides to proteolysis [50]. A common strategy to improve the stability is to produce the peptide as a fusion [147]. A major disadvantage with this strategy is that the desired product only constitutes a small portion of the fusion protein, often resulting in low yields of the target peptide. One way of increasing the molar ratio, and hence increase the amount of peptide produced, is to produce a fusion protein with multiple copies of the target peptide. An additional beneficial effect is often obtained by this strategy, since the gene multimerization has also been shown to increase the proteolytic stability of the produced peptides [173]. When the gene multimerization strategy is employed to increase the production yield, subsequent processing of the gene product to obtain the native peptide is needed. By flanking a peptide gene with codons encoding methionine, CNBr cleavage of the fusion protein, containing multiple repeats of the peptide, has successfully been used for obtaining native peptide at high yield [174,175]. Takasuga and co-workers [176] produced a pentapeptide multimerized to 3, 14 and 28 copies, fused to dihydrofolate reductase, engineered to be separated by trypsin cleavage. A similar strategy was used to produce a peptide hormone of 28 residues [177]. Eight copies of the peptide gene were linked in tandem, separated by codons specifying lysine residues flanking the peptide, and the construct was fused to a gene fragment encoding a portion of β-galactosidase. Endoproteinase Lys-C, an enzyme which specifically cleaves on the C-terminal side of lysine residues, was used instead of trypsin, together with carboxypeptidase B, to release the native peptide. Similarly, a multimerization strategy was used to improve the yields of the 31-amino-acid human proinsulin C-peptide [95]. The C-peptide was expressed intracellularly in *E. coli* as one, three or seven copies as parts of fusion proteins. Since it was found that the three different fusion protein were expressed at equal levels, and that they all were efficiently processed by trypsin/carboxypeptidase B treatment to release native C-peptide, the seven-copy construct was used to generate a recombinant production process [96].

Simplified site-specific removal of fusion partners

Genetically designed recombinant proteases have been used to simplify the removal of the proteases after site-specific cleavage of fusion proteins [3]. By fusing the protease to the same affinity tag as the target protein, an efficient removal of the affinity-tagged protease, the released affinity tag and uncleaved fusion protein can be achieved using affinity chromatography [156,157,168]. This principle is commercially available, examples being the systems based on His-tagged tobacco-etch-virus protease [158] and human rhinovirus 3C protease fused to a glutathione S-transferase tag (PreScission™ protease) [156]. An affinity-tagged protease can, as an alternative to covalent coupling, also be immobilized to an affinity matrix and be utilized for on-column cleavage [156]. On-column cleavage, in which the produced fusion proteins are site-specifically cleaved while still immobilized on the affinity column, has also been described [178,179]. An affinity-fusion system, consisting of a protein splicing intein domain from *S. cerevisiae* and a chitin-binding domain, allows simultaneous affinity purification and on-column cleavage [180,181]. Different immobilizing approaches are especially important for large-scale applications, since they can reduce the protease consumption and help to avoid additional contamination by the added protease.

Tailor-made product-specific affinity ligands

Powerful *in vitro* selection technologies, such as phage display [182], have proven efficient for the isolation of novel binding proteins from large collections (libraries) of peptides or proteins constructed, for example, by combinatorial protein engineering [183–186]. One example of such binding proteins is the so-called 'affibodies', selected from libraries constructed by random mutagenesis of the Z domain derived from SPA [187]. The Z domain, used as scaffold during library constructions, is proteolytically stable, highly soluble, small (6 kDa), and has a compact and robust structure devoid of intramolecular disulphide bridges, making it an ideal domain for ligand development. Using phage-display technology, affibody ligands to a wide range of targets have been successfully selected [188,189]. Recently, such affibody ligands showed selective binding in authentic affinity-chromatographic applications involving the purification of target proteins from *E. coli* total cell lysates [190]. Such tailor-made product-specific affinity ligands have also been generated and used for highly efficient recovery of recombinant human Factor VIII produced in Chinese-hamster ovary ('CHO') cells [191], and a recombinant vaccine can-

didate, derived from the RSV G protein, produced in baby-hamster kidney ('BHK') cells [192].

The obvious advantage of using a ligand selected to bind to the target protein instead of fusing the target protein to an affinity tag is that no cleavage step to obtain the native protein is needed. The disadvantage is that a new high-affinity ligand must be selected and produced for every new recombinant protein needed to be purified. It is nevertheless likely that this strategy will be attractive in recombinant bioprocesses, since highly selective affinity matrices can be created that potentially even could discriminate between different folding forms of the target protein and could thus replace several other chromatographic steps in the recovery process. Interestingly, no loss of column capacity or selectivity for the target protein was obtained even after repeated cycles of low pH elution and column sanitation protocols, including 0.5 M NaOH [190]. This might suggest that affinity chromatography using protein ligands could become increasingly used also in industrial-scale recombinant-proteins recovery processes in the future.

Concluding remarks

A number of different genetic strategies are available for the design of processes for production of recombinant proteins. In this review, the use of several such strategies in order to increase expression yields and simplify the recovery processes have been described. First of all, genetic strategies have been used for at least two decades to create efficient expression vectors in which suitable promoters are included and transcription and translation efficiencies are taken into account. Furthermore, the bacterial strains can be improved in different ways to give increased yields, for example by deletion of protease genes or engineered for overexpression of rare-codon tRNAs, foldases or chaperones. It has also been demonstrated that gene-multimerization strategies can be beneficial in improving production yields. All genetic strategies used to obtain increased expression yields are very much empirical and have to be performed on a product-by-product bases.

Another application of genetic strategies is the improvement of the efficiency of the recovery processes by adapting the product for a particular purification principle. In this area, affinity fusions have been commonly used, but also other principles, such as modified pI or hydrophobic properties, have been successfully investigated. A general drawback with fusion strategies is that the fusion partner often has to be removed to release a native gene product. A drastic step forward in the use of gene technology to improve recovery processes for recombinant proteins is the introduction of combinatorial protein engineering to generate tailor-made product-specific affinity ligands. This strategy, which allows highly specific recovery of a recombinant protein that can be expressed in its native form [188,191], is likely to be increasingly used also in industrial-scale bioprocesses, since this novel type of protein ligand (e.g. the protein A-based affibodies) can be sanitized using common industrial cleaning-in-place procedures [190].

Taken together, the examples presented make it evident that genetic strategies have had significant impact on recombinant-protein production during the last few decades, and that genetic design will be of utmost importance in the future for facilitating production and recovery of recombinant proteins.

References

1 Rudolph, R. (1996) in Protein Engineering: Principles and Practice (Cleland, J. L. and Craik, C. S., eds.), pp. 283–298, Wiley-Liss, New York
2 Cornelis, P. (2000) Curr. Opin. Biotechnol. **11**, 450–454
3 Nilsson, J., Ståhl, S., Lundeberg, J., Uhlén, M. and Nygren, P.-Å. (1997) Protein Expr. Purif. **11**, 1–16
4 Stevens, R. C. (2000) Structure **8**, R177–R185
5 Georgiou, G. (1996) in Protein engineering: Principles and Practice (Cleland, J. L. and Craik, C. S., eds.), pp. 101–127, Wiley-Liss, New York
6 Makrides, S. C. (1996) Microbiol. Rev. **60**, 512–538
7 Baneyx, F. (1999) Curr. Opin. Biotechnol. **10**, 411–421
8 Balbas, P. and Bolivar, F. (1990) Methods Enzymol. **185**, 14–37
9 Friehs, K. and Reardon, K. F. (1993) Adv. Biochem. Eng. Biotechnol. **48**, 53–77
10 Yansura, D. G. and Henner, D. J. (1990) Methods Enzymol. **185**, 54–60
11 Gronenborn, B. (1976) Mol. Gen. Genet. **148**, 243–250
12 Amann, E., Ochs, B. and Abel, K. J. (1988) Gene **69**, 301–315
13 de Boer, H. A., Comstock, L. J. and Vasser, M. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 21–25
14 Brosius, J., Erfle, M. and Storella, J. (1985) J. Biol. Chem. **260**, 3539–3541
15 Studier, F. W. and Moffatt, B. A. (1986) J. Mol. Biol. **189**, 113–130
16 Studier, F. W. (1991) J. Mol. Biol. **219**, 37–44
17 Guzman, L.-M., Belin, D., Carson, M. J. and Beckwith, J. (1995) J. Bacteriol. **177**, 4121–4130
18 Bernard, H. U., Remaut, E., Hershfield, M. V., Das, H. K., Helinski, D. R., Yanofsky, C. and Franklin, N. (1979) Gene **5**, 59–76
19 Nilsson, B. and Abrahmsén, L. (1990) Methods Enzymol. **185**, 144–161
20 Hasan, N. and Szybalski, W. (1995) Gene **163**, 35–40
21 Löfdahl, S., Guss, B., Uhlén, M., Philipson, L. and Lindberg, M. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 697–701

22 Hannig, G. and Makrides, S. C. (1998) Trends Biotechnol. **16**, 54–60
23 Neubauer, P., Hofmann, K., Holst, O., Mattiasson, B. and Kruschke, P. (1992) Appl. Microbiol. Biotechnol. **36**, 739–744
24 Yoon, S. K., Kang, W. K. and Park, T. H. (1996) J. Ferment. Bioeng. **81**, 153–157
25 Weickert, M. J., Doherty, D. H., Best, E. A. and Olins, P. O. (1996) Curr. Opin. Biotechnol. **7**, 494–499
26 Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D. and Higgins, C. F. (1987) Cell **48**, 297–310
27 Swartz, J. R. (2001) Curr. Opin. Biotechnol. **12**, 195–201
28 Brosius, J., Ullrich, A., Raker, M. A., Gray, A., Dull, T. J., Gutell, R. R. and Noller, H. F. (1981) Plasmid **6**, 112–118
29 Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D. and Gold, L. (1992) Mol. Microbiol. **6**, 1219–1229
30 Hall, M. N., Gabay, J., Debarbouille, M. and Schwartz, M. (1982) Nature (London) **295**, 616–618
31 Chen, H., Pomeroy-Cloney, L., Bjerknes, M., Tam, J. and Jay, E. (1994) J. Mol. Biol. **240**, 20–27
32 Stenström, C. M., Jin, H., Major, L. L., Tate, W. P. and Isaksson, L. A. (2001) Gene **263**, 273–284
33 de Boer, H. A. and Kastelein, R. A. (1986) in Maximizing Gene Expression (Reznikoff, W. and Gold, L., eds.), pp. 225, Butterworths, Boston
34 Bula, C. and Wilcox, K. W. (1996) Protein Expression Purif. **7**, 92–103
35 Kane, J. F. (1995) Curr. Opin. Biotechnol. **6**, 494–500
36 Zahn, K. (1996) J. Bacteriol. **178**, 2926–2933
37 Forman, M. D., Stack, R. F., Masters, P. S., Hauer, C. R., Baxter, S. M. (1998) Protein Sci. **7**, 500–503
38 Sharp, P. M. and Bulmer, M. (1988) Gene **63**, 141–145
39 Poole, E. S., Brown, C. M. and Tate, W. P. (1995) EMBO J. **14**, 151–158
40 Lee, S. Y. (1996) Trends Biotechnol. **14**, 98–105
41 Kleman, G. L., Chalmers, J. J., Luli, G. W. and Strohl, W. R. (1991) Appl. Environ. Microbiol. **57**, 910–917
42 Yee, L. and Blanch, H. W. (1992) Bio/Technology **10**, 1550–1556
43 Kleman, G. L. and Strohl, W. R. (1994) Curr. Opin. Biotechnol. **5**, 180–186
44 Riesenberg, D., Schulz, V., Knorre, W. A., Pohl, H. D., Korz, D., Sanders, E. A., Ross, A. and Deckwer, W. D. (1991) J. Biotechnol. **20**, 17–27
45 Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A. and Deckwer, W. D. (1995) J. Biotechnol. **39**, 59–65
46 Aristidou, A. A., San, K. Y. and Bennett, G. N. (1999) Biotechnol. Prog. **15**, 140–145
47 Han, K., Hong, J. and Lim, H. C. (1993) Biotechnol. Bioeng. **41**, 316–324
48 Hahm, D. H., Pan, J. and Rhee, J. S. (1994) Appl. Microbiol. Biotechnol. **42**, 100–107
49 Aristidou, A. A., San, K. Y. and Bennett, G. N. (1995) Biotechnol. Prog. **11**, 475–478
50 Murby, M., Uhlén, M. and Ståhl, S. (1996) Protein Expression Purif. **7**, 129–136
51 Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F. et al. (1997) Science **277**, 1453–1474
52 Hansson, M., Ståhl, S., Hjorth, R., Uhlén, M. and Moks, T. (1994) Bio/Technology **12**, 285–288
53 Sandkvist, M. and Bagdasarian, M. (1996) Curr. Opin. Biotechnol. **7**, 505–511
54 Pugsley, A. P., Francetic, O., Possot, O. M., Sauvonnet, N. and Hardie, K. R. (1997) Gene **192**, 13–19
55 Ståhl, S., Nygren, P.-Å. and Uhlén, M. (1997) in Methods in Molecular Biology, Recombinant Protein Protocols: Detection and Isolation (Tuan, R., ed.), pp. 103–118, Humana Press Inc., Totowa
56 Neri, D., de Lalla, C., Petrul, H., Neri, P. and Winter, G. (1995) Bio/Technology **13**, 373–377
57 Ko, J. H., Park, D. K., Kim, I. C., Lee, S. H. and Byun, S. M. (1995) Biotechnol. Lett. **17**, 1019–1024
58 Hancock, R. E. W. (1983) Annu. Rev. Microbiol. **38**, 237–264
59 van der Wal, F. J., Koningstein, G., ten Hagen, C. M., Oudega, B. and Luirink, J. (1998) Appl. Environ. Microbiol. **64**, 392–398
60 Wan, E. W. and Baneyx, F. (1998) Protein Expression Purif. **14**, 13–22
61 Ariga, O., Andoh, Y., Fujishita, Y., Watari, T. and Sano, Y. (1991) J. Ferment. Bioeng. **71**, 397–402
62 Aristidou, A. A., Yu, P. and San, K.-Y. (1993) Biotechnol. Lett. **15**, 331–336
63 Blight, M. A., Chervaux, C. and Holland, I. B. (1994) Curr. Opin. Biotechnol. **5**, 468–474
64 Kadonaga, J. T., Gautier, A. E., Straus, D. R., Charles, A. D., Edge, M. D. and Knowles, J. R. (1984) J. Biol. Chem. **259**, 2149–2154
65 Collins-Racie, L. A., McColgan, J. M., Grant, K. L., DiBlasio-Smith, E. A., McCoy, J. M. and LaVallie, E. R. (1995) Bio/Technology **13**, 982–987
66 Meerman, H. J. and Georgiou, G. (1994) Bio/Technology **12**, 1107–1110
67 Goldberg, A. L. (1992) Eur. J. Biochem. **203**, 9–23
68 Maurizi, M. R. (1992) Experientia **48**, 178–201
69 Enfors, S. O. (1992) Trends Biotechnol. **10**, 310–315
70 Baneyx, F. and Georgiou, G. (1992) Ann. N.Y. Acad. Sci. **665**, 301–308
71 Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. and Uhlén, M. (1987) Protein Eng. **1**, 107–113
72 Nygren, P.-Å., Eliasson, M., Palmcrantz, E., Abrahmsén, L. and Uhlén, M. (1988) J. Mol. Recognit. **1**, 69–74
73 Hammarberg, B., Nygren, P.-Å., Holmgren, E., Elmblad, A., Tally, M., Hellman, U., Moks, T. and Uhlén, M. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 4367–4371

74 Murby, M., Cedergren, L., Nilsson, J., Nygren, P.-Å., Hammarberg, B., Nilsson, B., Enfors, S. O. and Uhlén, M. (1991) Biotechnol. Appl. Biochem. **14**, 336–346

75 Wülfing, C. and Plückthun, A. (1994) Mol. Microbiol. **12**, 685–692

76 Missiakas, D., Schwager, F. and Raina, S. (1995) EMBO J. **14**, 3415–3424

77 Wall, J. G. and Plückthun, A. (1995) Curr. Opin. Biotechnol. **6**, 507–516

78 Georgiou, G. and Valax, P. (1996) Curr. Opin. Biotechnol. **7**, 190–197

79 Sone, M., Kishigami, S., Yoshihisa, T. and Ito, K. (1997) J. Biol. Chem. **272**, 6174–6178

80 Thomas, J. G., Ayling, A. and Baneyx, F. (1997) Appl. Biochem. Biotechnol. **66**, 197–238

81 Hober, S. and Uhlén, M. (1999) in Protein Engineering in Industrial Biotechnology (Alberghina, L., ed.), pp. 23–39, Harwood Academic Publishers, Amsterdam

82 Bothmann, H. and Plückthun, A. (2000) J. Biol. Chem. **275**, 17100–17105

83 Samuelsson, E., Jonasson, P., Viklund, F., Nilsson, B. and Uhlén, M. (1996) Nat. Biotechnol. **14**, 751–755

84 Lilie, H., Schwarz, E. and Rudolph, R. (1998) Curr. Opin. Biotechnol. **9**, 497–501

85 Schein, C. H. (1993) Curr. Opin. Biotechnol. **4**, 456–461

86 Murby, M., Samuelsson, E., Nguyen, T. N., Mignard, L., Power, U., Binz, H., Uhlén, M. and Ståhl, S. (1995) Eur. J. Biochem. **230**, 38–44

87 Dale, G. E., Broger, C., Langen, H., D'Arcy, A. and Stuber, D. (1994) Protein Eng. **7**, 933–939

88 LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. (1993) Bio/Technology **11**, 187–193

89 Wilkinson, D. L., Ma, N. T., Haught, C. and Harrison, R. G. (1995) Biotechnol. Prog. **11**, 265–269

90 Baker, R. T. (1996) Curr. Opin. Biotechnol. **7**, 541–546

91 Pilon, A. L., Yost, P., Chase, T. E., Lohnas, G. L. and Bentley, W. E. (1996) Biotechnol. Prog. **12**, 331–337

92 Davies, G. D., Elisee, C., Newham, D. M. and Harrison, R. G. (1999) Biotechnol. Bioeng. **65**, 382–388

93 Liao, H. H. (1991) Protein Expression Purif. **2**, 43–50

94 Larsson, M., Brundell, E., Nordfors, L., Höög, C., Uhlén, M. and Ståhl, S. (1996) Protein Expression Purif. **7**, 447–457

95 Jonasson, P., Nygren, P.-Å., Johansson, B.-L., Wahren, J., Uhlén, M. and Ståhl, S. (1998) Gene **210**, 203–210

96 Jonasson, P., Nygren, P.-Å., Jörnvall, H., Johansson, B.-L., Wahren, J., Uhlén, M. and Ståhl, S. (2000) J. Biotechnol. **76**, 215–226

97 di Guan, C., Li, P., Riggs, P. D. and Inouye, H. (1988) Gene **67**, 21–30

98 Kapust, R. B. and Waugh, D. S. (1999) Protein Sci. **8**, 1668–1674

99 Zhang, Y., Olsen, D. R., Nguyen, K. B., Olson, P. S., Rhodes, E. T. and Mascarenhas, D. (1998) Protein Expression Purif. **12**, 159–165

100 Hockney, R. C. (1994) Trends Biotechnol. **12**, 456–463

101 Qiu, J. Swartz, J. R. and Georgiou, G. (1998) Appl. Environ. Microbiol. **64**, 4891–4896

102 Nishihara, K., Kanemori, M., Yanagi, H. and Yura, T. (2000) Appl. Environ. Microbiol. **66**, 884–889

103 Derman, A. I., Prinz, W. A., Belin, D. and Beckwith, J. (1993) Science **262**, 1744–1747

104 Proba, K., Ge, L. and Plückthun, A. (1995) Gene **159**, 203–207

105 Bessette, P. H., Åslund, F., Beckwith, J. and Georgiou, G. (1999) Proc. Natl. Acad. Sci. U.S.A. **96**, 13703–13708

106 Harrison, S. T., Dennis, J. S. and Chase, H. A. (1991) Bioseparation **2**, 95–105

107 Foster, D. (1995) Curr. Opin. Technol. **6**, 523–526

108 Lapidot, A., Mechaly, A. and Shoham, Y. (1996) J. Biotechnol. **51**, 259–264

109 Rüdiger, A., Jorgensen, P. L. and Antranikian, G. (1995) Appl. Environ. Microbiol. **61**, 567–575

110 Mackey, B. M., Miles, C. A., Parsons, S. E. and Seymour, D. A. (1991) J. Gen. Microbiol. **137**, 2361–2374

111 Collis, M. A., O'Neill, B. K., Thomas, C. J. and Middelberg, A. P. (1996) Bioseparation **6**, 55–63

112 Watson, J. S. and Cumming, R. H. (1987) in Proceedings of the Fourth European Congress of Biotechnology (Neijssel, O., Meer, R. and Luyben, K., eds.), pp. 526–529, Elsevier Science, Amsterdam

113 Ford, C. F., Suominen, I. and Glatz, C. E. (1991) Protein Expression Purif. **2**, 95–107

114 Strandberg, L. and Enfors, S. O. (1991) Appl. Environ. Microbiol. **57**, 1669–1674

115 Valax, P. and Georgiou, G. (1993) Biotechnol. Prog. **9**, 539–547

116 Clark, E. D. B. (1998) Curr. Opin. Biotechnol. **9**, 157–163

117 Rudolph, R., Böhm, G., Lilie, H. and Jaenicke, R. (1997) in Protein Function, a Practical Approach (Creighton, T. E., ed.), pp. 57–99, IRL Press, Oxford

118 Rudolph, R. and Lilie, H. (1996) FASEB J. **10**, 49–56

119 Hevehan, D. and Clark, D. B. C. (1997) Biotechnol. Bioeng. **54**, 221–230

120 Altamirano, M. M., García, C., Possani, L. D. and Fersht, A. R. (1999) Nat. Biotechnol. **17**, 187–191

121 Altamirano, M. M., Woolfson, A., Donda, A., Shamshiev, A., Briseño-Roa, L., Foster, N. W., Veprintsev, D. B., De Libero, G., Fersht, A. R. and Milstein, C. (2000) Proc. Natl. Acad. Sci. U.S.A. **98**, 3288–3293

122 Carlson, J. D. and Yarmush, M. L. (1992) Bio/Technology **10**, 86–91

123 Katzav-Gozansky, T., Hanan, E. and Solomon, B. (1996) Biotechnol. Appl. Biochem. **23**, 227–230

124 Hober, S., Hansson, A., Uhlén, M. and Nilsson, B. (1994) Biochemistry **33**, 6758–6761

125 Cowley, D. J. and Mackin, R. B. (1997) FEBS Lett. **402**, 124–130
126 Jonasson, P., Nilsson, J., Samuelsson, E., Moks, T., Ståhl, S. and Uhlén, M. (1996) Eur. J. Biochem. **236**, 656–661
127 Nilsson, J., Jonasson, P., Samuelsson, E., Ståhl, S. and Uhlén, M. (1996) J. Biotechnol. **48**, 241–250
128 Hjorth, R. (1997) Trends Biotechnol. **15**, 230–235
129 Barnfield Frej, A.-K., Johansson, H. J., Johansson, S. and Leijon, P. (1997) Bioprocess Eng. **16**, 57–63
130 Glatz, C. E. (1990) Bioprocess Technol. **9**, 329–356
131 Gupta, M. N., Kaul, R., Guoqiang, D., Dissing, U. and Mattiasson, B. (1996) J. Mol. Recognit. **9**, 356–359
132 Tjerneld, F. and Johansson, G. (1990) Bioseparation **1**, 255–263
133 Huddleston, J., Veide, A., Köhler, K., Flanagan, J., Enfors, S. O. and Lyddiatt, A. (1991) Trends Biotechnol. **9**, 381–388
134 Kamihira, M. and Mattiasson, B. (1992) Biotechnol. Bioeng. **40**, 1382–1387
135 Freitag, R. and Horvath, C. (1996) Adv. Biochem. Eng. Biotechnol. **53**, 17–59
136 Sofer, G. and Hagel, L. (1997) Handbook of Process Chromatography: A Guide to Optimization, Scale-Up, and Validation, Academic Press, San Diego
137 Ersson, B., Rydén, L. and Jansson, J.-C. (1998) in Protein Purification: Principles, High-Resolution Methods, and Applications (Jansson, J.-C. and Rydén, L., eds.), pp. 3–40, John Wiley and Sons, New York
138 Uhlén, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. (1983) Gene **23**, 369–378
139 Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975) Nature (London) **258**, 598–599
140 Smith, D. B. and Johnson, K. S. (1988) Gene **67**, 31–40
141 Hopp, T. H., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L. and Conlon, P. J. (1988) Bio/Technology **6**, 1204–1210
142 Brizzard, B. L., Chubet, R. G. and Vizard, D. L. (1994) Bio Techniques **16**, 730–735
143 Cronan, J. E. (1990) J. Biol. Chem. **265**, 10327–10333
144 Schatz, P. J. (1993) Bio/Technology **11**, 1138–1143
145 Porath, J. (1992) Protein Expression Purif. **3**, 263–281
146 Nygren, P.-Å., Ståhl, S. and Uhlén, M. (1994) Trends Biotechnol. **12**, 184–188
147 Ståhl, S., Nilsson, J., Hober, S., Uhlén, M. and Nygren, P.-Å. (1999) in The Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis and Bioseparation (Flickinger, M. C. and Drew, S. W., eds.), pp. 49–63, John Wiley and Sons, New York
148 Sofer, G. (1995) J. Chromatogr. A **707**, 23–28
149 La Vallie, E. R., McCoy, J. M., Smith, D. B. and Riggs, P. (1994) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 16.4.5–16.4.17, John Wiley and Sons, New York
150 Nilsson, B., Forsberg, G. and Hartmanis, M. (1991) Methods Enzymol. **198**, 3–16
151 Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H. W. (1977) Science **198**, 1056–1063
152 Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K. M. and Hartmanis, M. (1991) J. Protein. Chem. **10**, 517–526
153 Pohlner, J., Kramer, J. and Meyer, T. F. (1993) Gene **130**, 121–126
154 Maina, C. V., Riggs, P. D., Grandea, A. G.d., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A. and Guan, C. D. (1988) Gene **74**, 365–373
155 Chang, J.-Y. (1985) Eur. J. Biochem. **151**, 217–224
156 Walker, P. A., Leong, L. E., Ng, P. W., Tan, S. H., Waller, S., Murphy, D. and Porter, A. G. (1994) Bio/Technology **12**, 601–605
157 Gräslund, T., Nilsson, J., Lindberg, A. M., Uhlén, M. and Nygren, P.-Å. (1997) Protein Expression Purif. **9**, 125–132
158 Parks, T. D., Leuther, K. K., Howard, E. D., Johnson, S. A. and Dougherty, W. G. (1994) Anal. Biochem. **216**, 413–417
159 Carter, P. (1990) in Protein Purification: From Molecular Mechanisms to Large-Scale Processes (Ladish, M. R., Willson, C.-D., Painton, C.-D. C. and Builder, S. E., eds.), pp. 181–193, American Chemical Society, Washington
160 Köhler, K., Ljungquist, C., Kondo, A., Veide, A. and Nilsson, B. (1991) Bio/Technology **9**, 642–646
161 Carlsson, M., Berggren, K., Linse, P., Veide, A. and Tjerneld, F. (1996) J. Chromatogr. A **756**, 107–117
162 Zhao, J. Y., Ford, C. F., Glatz, C. E., Rougvie, M. A. and Gendel, S. M. (1990) J. Biotechnol. **14**, 273–283
163 Smith, J. C., Derbyshire, R. B., Cook, E., Dunthorne, L., Viney, J., Brewer, S. J., Sassenfeld, H. M. and Bell, L. D. (1984) Gene **32**, 321–327
164 Dalbøge, H., Dahl, H.-H. M., Pedersen, J., Hansen, J. W. and Christensen, T. (1987) Bio/Technology **5**, 161–164
165 Persson, M., Bergstrand, M. G., Bulow, L. and Mosbach, K. (1988) Anal. Biochem. **172**, 330–337
166 Gräslund, T., Lundin, G., Uhlén, M., Nygren, P.-Å. and Hober, S. (2000) Protein Eng. **13**, 703–709
167 Gräslund, T., Ehn, M., Lundin, G., Hedhammar, M., Uhlén, M., Nygren, P.-Å. and Hober, S. (2001) J. Chromatogr. A **942**, 157–166
168 Gräslund, T., Hedhammar, M., Nygren, P.-Å., Uhlén, M. and Hober, S. (2002) J. Biotechnol., in the press
169 Samuelsson, E., Wadensten, H., Hartmanis, M., Moks, T. and Uhlén, M. (1991) Bio/Technology **9**, 363–366
170 Samuelsson, E., Moks, T., Nilsson, B. and Uhlén, M. (1994) Biochemistry **33**, 4207–4211
171 Zouhar, J., Nanak, E. and Brzobohaty, B. (1999) Protein Expression Purif. **17**, 153–162
172 Stempfer, G., Holl-Neugebauer, B. and Rudolph, R. (1996) Nat. Biotechnol. **14**, 329–334
173 Shen, S.-H. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4627–4631

174 Kempe, T., Kent, S. B., Chow, F., Peterson, S. M., Sundquist, W. I., L'Italien, J. J., Harbrecht, D., Plunkett, D. and DeLorbe, W. J. (1985) Gene **39**, 239–245
175 Lee, J. H., Minn, I., Park, C. B. and Kim, S. C. (1998) Protein Expression Purif. **12**, 53–60
176 Takasuga, A., Banba, K., Yoshino, K., Izutsu, H., Iwakura, M. and Ohashi, S. (1992) J. Biochem. (Tokyo) **112**, 652–657
177 Lennick, M., Haynes, J. R. and Shen, S. H. (1987) Gene **61**, 103–112
178 Altman, J. D., Henner, D., Nilsson, B., Anderson, S. and Kuntz, I. D. (1991) Protein Eng. **4**, 593–600
179 Beswick, V., Brodsky, J. L., Kepes, F., Neumann, J. M., Sanson, A. and Garrigos, M. (1998) Protein Expression Purif. **13**, 423–432
180 Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B. and Xu, M. Q. (1996) J. Biol. Chem. **271**, 22159–22168
181 Xu, M. Q. and Perler, F. B. (1996) EMBO J. **15**, 5146–5153
182 Wilson, D. R. and Finlay, B. B. (1998) Can. J. Microbiol. **44**, 313–329
183 Clackson, T. and Wells, J. A. (1994) Trends Biotechnol. **12**, 173–184
184 Nygren, P.-Å. and Uhlén, M. (1997) Curr. Opin. Struct. Biol. **7** 463–469
185 Skerra, A. (2000) J. Mol. Recognit. **4**, 167–187
186 Ladner, R. C. and Ley, A. C. (2001) Curr. Opin. Biotechnol. **12**, 406–410
187 Nord, K., Nilsson, J., Nilsson, B., Uhlén, M. and Nygren, P.-Å. (1995) Protein Eng. **8**, 601–608
188 Nord, K., Gunneriusson, E., Ringdahl, J., Ståhl, S., Uhlén, M. and Nygren, P.-Å. (1997) Nat. Biotechnol. **15**, 772–777
189 Hansson, M., Ringdahl, J., Robert, A., Power, U., Goetsch, L., Nguyen, T. N., Uhlén, M., Ståhl, S. and Nygren, P.-Å. (1999) Immunotechnology **4**, 237–252
190 Nord, K., Gunneriusson, E., Uhlén, M. and Nygren, P.-Å. (2000) J. Biotechnol. **80**, 45–54
191 Nord, K., Nord, O., Uhlén, M., Kelley, B., Ljungqvist, C. and Nygren, P.-Å. (2001) Eur. J. Biotechnol. **268**, 4269–4277
192 Andersson, C., Hansson, M., Power, U., Nygren, P.-Å. and Ståhl, S. (2001) Biotechnol. Appl. Biochem. **34**, 25–32

Received 1 November 2001; accepted 18 November 2001