IMPROVEMENT OF DOWNSTREAM PROCESSING OF RECOMBINANT PROTEINS BY MEANS OF GENETIC ENGINEERING METHODS

ERWIN FLASCHEL and KARL FRIEHS

Universität Bielefeld, Technische Fakultät, Arbeitsgruppe Fermentationstechnik,
Postfach 10 01 31, 4800 Bielefeld 1, Germany

ABSTRACT

The rapid advancement of genetic engineering has allowed to produce an impressive number of proteins on a scale which would not have been achieved by classical biotechnology. At the beginning of this development research was focussed on elucidating the mechanisms of protein overexpression. The appearance of inclusion bodies may illustrate the success. In the meantime, genetic engineering is not only expected to achieve overexpression, but to improve the whole process of protein production. For downstream processing of recombinant proteins, the synthesis of fusion proteins is of primary importance. Fusion with certain proteins or peptides may protect the target protein from proteolytic degradation and may alter its solubility. Intracellular proteins may be translocated by means of fusions with signal peptides. Affinity tags as fusion complements may render protein separation and purification highly selective. These methods as well as similar ones for improving the downstream processing of proteins will be discussed on the basis of recent literature.

KEY WORDS

Downstream processing, recombinant protein, genetic engineering, fusion protein, fusion complement, purification tag, affinity handle, protein secretion, protein export.
Proteins are the primary products of the translation of the genetic code. They are therefore of central interest for life science research as well as applications in a multitude of areas. Thus, they are a very important class of biotechnology-derived products. Enzymes have found a large area of applications due to their catalytic properties. Spectacular advancements in molecular genetics (regulation of transcription and translation, gene transfer, polymerase chain reaction etc.) and in the cultivation of animal and human cells (cell fusion, hybridoma-technique, transfection) have given the opportunity that intensive efforts are undertaken in order to develop processes for the production of species-specific peptide hormones (insulin, somatostatin, growth hormones, etc.), immunologically active proteins (interferons, interleukins, etc.), coagulation factors (urokinase, t-PA, factor VIII, etc.), as well as of vaccines and monoclonal antibodies.

Such proteins may only be obtained economically by means of biotechnological processes, since they represent polymers composed out of the 21 different proteinogeneous L-α-amino acids (including selenocysteine). Chemical synthesis of proteins [1] is only competitive for the production of short chain peptides. However, chemical synthesis is important when non-proteinogeneous amino acids have to be incorporated into peptides. The chemical route suffers from a laborious and expensive protection group chemistry and an error rate which is still too high. In addition, it has to be taken into account that proteins are only active in their natural folding state and that proteins of eukaryotes often need posttranslational modifications (glycosylation, proteolysis). In consequence, protein synthesis will need the use of microorganisms and of cells of animal or human origin. This also means that proteins will need to be separated from complex natural media. Cell-free translation systems are another possible method to produce recombinant proteins. The usefulness of such systems in bioprocessing is under investigation [274]. The importance of downstream processing very likely will at least remain or even increase [2, 3] because the separation and purification of proteins will remain the main source of costs for the manufacturing of proteins.

A whole spectrum of downstream processes has been developed for the recovery of proteins [4-7] and these have been documented in recent books and monographs [8-15]. Uncertainty with the use of novel processes, the partiality for problem-specific solutions and new possibilities for the production of a multitude of pharmacologically active proteins have considerably promoted the interest in downstream processing. Since the tool box of genetic engineering is more and more used for improving the production of proteins, it has been straightforward not only to use these techniques in order to achieve overexpression but to aim at the improvement of downstream processing as well. Quite some novel approaches for improving the downstream processing of proteins have already been worked out [16-19]. These novel methods may lead to the opportunity that very different proteins might be
purified by means of one unique process strategy. On the one hand, the physico-chemical properties of the proteins may be changed or amplified by means of protein design or site-directed mutagenesis. Recombination of different genes for the production of fusion proteins, on the other hand, is of primary importance because unique fusion partners may be chosen for the complementation of the target protein. Such fusion partners are adequate which are carriers of a specific affinity useful for the downstream process applied. The class of fusion proteins englobe the use of signal sequences in order to change the localization of the target proteins.

Such approaches and similar ones will be reviewed - taken from recent literature. The following chapter will deal with questions referring to the choice of host organisms and the requirement for producing biologically active natural proteins. Only a few issues can be emphasized with respect to host organisms and expression in general for not overloading the article. A further chapter discusses possibilities of manipulating the localization of proteins and the properties of cells which have direct impact on downstream processing. The last chapter is dedicated to strategies of protein modification for the improvement of protein separation and purification.

**PRODUCT CONCENTRATION AND BIOLOGICAL ACTIVITY**

The economy of downstream processing of proteins highly depends on the achievable product concentration. Therefore, recombinant protein production aims at the improvement of the overall protein concentration. However, aiming at a high product concentration may fail, unless a high proportion of biologically active proteins is obtained.

**Overexpression**

The first step for improving the concentration of the target protein is the choice of an adequate biological system by means of which overexpression may be achieved. Many such expression systems are available. For the optimisation of these systems a series of factors have to be taken into account referring to the replication, transcription and translation of the genes.

The replication of genes coding for recombinant proteins strongly influences the stable transmission as well as the copy number of these genes. Most expression systems for bacteria use plasmids as vehicles of recombinant DNA. Plasmid stability depends on both the structural and the segregative stability. The structural stability refers to the error rate during plasmid replication, whereas the segregative stability depends on the transmission of the plasmid on to daughter cells. Besides the correct replication of plasmids the structural stability is influenced by the rate of mutation and the gene repair facilities of the host organism. The segregative stability, commonly called plasmid stability, depends on a
series of factors which cannot be discussed here in detail [20]. Means for stabilizing or enhancing plasmid stability have been discussed by Kumar et al. [21].

The rate of gene expression may be accelerated by increasing the number of plasmids per cell. This may be achieved by using so-called multi-copy plasmids, the copy number of which may reach 100 owing to their particular mode of replication. Since the plasmid copy number has to be high only at the time when the target protein shall be produced, a controlled induction of plasmid replication represents an advantageous strategy. This may be achieved by using so-called runaway replication plasmids the copy number of which may easily reach 1,000 [22]. An obvious additional strategy consists of integrating several copies of the coding gene into the plasmid. However, it has to be kept in mind that a high plasmid copy number may affect the physiological state of the host organism leading to a limited rate of gene expression.

Both a promoter-operator system and an appropriate termination sequence are required in order to achieve efficient transcription. In addition, transcription should be inducible. During the development of expression systems different sequences have been recognized as strong promoters. A strong promoter is the prerequisite of a high transcription frequency. The best known inducible bacterial promoter-operator systems are those using operators derived from the lac- [23], trp- [24] and the λ- [25] systems. The lac-operator is often used for the induction of different promoters like lac and tac. Induction is achieved by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside). Promoters coupled with the trp-operator may be induced by IAA (3-(3-indolyi)-acrylic acid). An induction by means of a temperature jump is applied in the case of λ-promoters using a temperature sensitive mutant strain with respect to the λ-repressor. Chemical inducers may be quite expensive and may lead to contamination of the product. Temperature induction may cause undesirable secondary products to appear due to an altered metabolism. Therefore, there is still a need for improved novel promoters [281] - like ones inducible by oxygen depletion [26] or by means of a pH-shift [279]. When strong promoters are applied, a strong terminator is required as well. Without an appropriate termination sequence transcription would proceed beyond the target gene and transcription would lead to useless products. Appropriate sequences like the trpA-transcription terminator are commercially available [e.g. 27].

Transcription is the first step during gene expression leading to mRNA. The second step involves the ribosomes which are responsible for the translation of the mRNA into the amino acid sequence of the recombinant protein. Translation efficiency depends on the stability of the mRNA, the ribosomal binding site (RBS), the correct termination of translation as well as the use of special codons. The stability of mRNA is a function of its susceptibility for hydrolysis by ribonucleases. It is commonly assumed that the lifetime of mRNA is influenced mainly by the presence of 3'-ribonucleases, because the secondary
structure of the 3'-terminal sequence of the mRNA is of crucial importance [28]. In consequence, mRNA stability may be enhanced by changing the 3'-terminal sequence or shifts in growth rate [264]. The lifetime of the mRNA determines how often it may be used by the ribosomes for translation [29]. Newest results suggest that RNA degradation may depend on 5'-terminal base pairing in *E. coli* [221].

Recombinant genes are commonly introduced behind an inducible promoter, but they often have their own ribosomal binding site (RBS). This may be the reason for a low translation frequency, although the rate of transcription may be high. Optimized gene expression systems include an RBS especially designed for the particular host organism [30]. Thus, AT-rich sequences may be introduced up- and downstream of the RBS in order to increase the initiation frequency of translation [31]. The distance between RBS and the start codon AUG may also be changed in order to achieve higher productivities [32]. Synthetic RBS-sequences have not led to improvements of gene expression [33]. A strong terminator is required for efficient translation, too. Appropriate sequences are commercially available [27].

If a synthetic gene is used the sequence of which has been deduced from the amino acid sequence of a protein, the base sequences of the codons for the amino acids should be chosen with care. Studies with genes of highly expressed proteins have shown that certain codons out of the pool of the degenerate genetic code are preferred [34, 266].

**Protein Stability**

After translation the stability of the protein itself determines how much product will be obtained. Protein stability mainly depends on the susceptibility of the protein for proteolytic decomposition. Especially in the case of recombinant proteins, proteases may cause a considerable loss of product [35, 222]. It seems as if heterologous proteins were recognized as belonging to a different species with the consequence of a rapid proteolytic degradation. Different proteases may be responsible for protein digestion. In the case of *E. coli* the protease La coded on the *lon*-gene is accused of mainly being responsible for the degradation of recombinant proteins [36]. By using *lon*-minus mutants this problem may be partially overcome. However, the use of such mutants may cause a series of difficulties during cultivation. The product of the *lon*-gene is also controlling cell division and the development of the polysaccharide capsule. Therefore, *lon*-minus mutants tend to overproduce polysaccharides causing strong production of slime and interference with cell division and growth [35]. This problem may be circumvented by controlling the *lon*-gene by inducing the *lon*-function during the growth phase and switching it off during the production phase. However, this control is quite delicate since the concentration of the *lon*-gene product has to be kept in a very narrow band. This may not be achieved with commonly used promoters because of the constitutive gene expression taking place in spite of
repression [37]. By inactivation of structural and regulatory genes of the slime capsule synthesis it is possible to prevent slime production and the growth problems related with that [38]. Modifications of the sulA-gene the product of which regulates the Ion-coded protease may yield a significant improvement of cell division for lon-negative strains. However, it has to be taken into account that lon-negative strains are very sensitive with respect to the induction of stress proteins causing additional growth problems.

Additional methods to reduce protein degradation are to inhibit the host-mediated degradation of foreign proteins by phage infection, using T4-promoter [262] and to consist the influence of growth rate on proteolysis [271]. An easier to implement strategy for reducing protein degradation consists in the synthesis of fusion proteins. Thus, small heterologous peptides which commonly are very labile have been stabilized by fusion with large proteins like ß-galactosidase or Staphylococcus protein A [39-43, 272]. In the case of such fusions it may be necessary to split off the fusion complement in order to obtain the target protein in the desired form.

The export of proteins is another elegant possibility to avoid the intracellular environment rich of proteases [44]. However, some secretion systems may also export proteases into the extracellular environment. The problems related to the presence of proteases have to be considered especially for the operation of cell disintegration [45]. Therefore, the addition of protease inhibitors may be necessary during cell disruption.

**Posttranslational Modification**

Many proteins, in particular eukaryotic proteins of pharmaceutical interest, require further modification after protein biosynthesis. Such posttranslational modifications may consist in phosphorylation, glycosylation, the cleavage of sequences from pre-pro-proteins of especially secretory proteins as well as the cleavage of the primary protein chain. Without appropriate modification these proteins are often obtained in a biologically inactive form.

In many cases it is relatively easy to produce large quantities of protein by means of bacterial systems like *E. coli*. However, strains of *E. coli* are not provided with a pathway for the glycosylation of proteins. If glycosylation is of crucial importance, other less convenient expression systems have to be used. Yeast cells, simple eukaryotes, have glycosylating activity, but often they use other glycosidic residues then those required for the particular case [46]. There is certainly still a long way to go until bacterial systems or yeasts may have learned to achieve different posttranslational modifications by means of genetic engineering. Cells of animal and human origin will remain the systems of choice for proteins needing adequate posttranslational processing [4]. Another way to circumvent this problem is to look for the biologically active part of the target protein which, by circumstance, may not need further modification [47].
Recombinant proteins are often found as insoluble aggregates in the cytoplasm [17, 48]. Extremely high protein concentrations due to overexpression may be responsible for this phenomenon [270]. These accumulations of solid insoluble proteins are called inclusion bodies [49]. The formation of these refractile particles may represent an advantage as well as a disadvantage for the production of recombinant proteins. Thus, inclusion bodies are relatively simple to be recovered from cell homogenates [17, 50, 51], and are protected from proteolytic cleavage. However, a great disadvantage must be seen in the fact that proteins bound in inclusion bodies are biologically inactive. Therefore, these highly insoluble protein aggregates have first to be denatured under extreme conditions in order to be solubilized [51, 52]. Subsequently, these solubilized proteins have to be renatured by means of adequate methods in order to be transformed into their biologically active form. This procedure may yield heavy losses which may be acceptable if a high overproduction goes in common with a high added value. Genetic engineering methods for the reduction of aggregate formation consist in forcing protein secretion as well as in changing specific properties of the target protein which may be responsible for the occurrence of inclusion bodies. Wilkinson et al. [48] describe factors influencing the solubility of recombinant proteins. The critical factors englobe the protein concentration, the distribution of surface charge, the portion of particular secondary structures, the contents of cysteine and proline as well as the overall hydrophobicity. These properties may be changed deliberately by site-directed mutagenesis as well as by fusion of the target protein with an appropriate complement [53, 277]. Some approaches along these lines have been discussed by Kotewicz [54]. The most simple possibility for reducing or even preventing inclusion body formation seems to be by decreasing the rate of expression [52, 223], e.g. by lowering the temperature after induction [224].

In the case of heterologous proteins it is often inevitable that aggregates are formed and the whole procedure of solubilization and renaturation has to be applied [55]. The achievement of the native structure is the most crucial aspect for the production of recombinant proteins [46]. Achieving correct protein folding is a fundamental problem and of enormous economic impact [51]. Protein folding depends on a variety of different factors [56-58]. Certain additives like e.g. PEG [59] may improve folding [60]. However, the kinetics of folding reveals a rather complex pathway [225].

In recent years it has become clear that in vivo protein folding is assisted by catalytic protein complexes. Such proteins obviously play a crucial role for protein folding and assembly. They belong to the heterogeneous group of heat-shock proteins (hsp), and are now divided into two more or less distinct classes [226]. The class of the chaperonins belonging to the hsp10 and hsp60 families occur both in eubacteria, mitochondria and plastids, whereas the class of the hsp70 family seems to be present everywhere where
protein folding occurs. Evidence is also emerging for the presence of cytosolic chaperonins in archaeabacteria as well as in eukaryotes [226, 228]. However, the term chaperone is generally used for proteins which assist in protein folding. Anyway, the division of chaperones into two classes seems to be a quite simplistic approach, because it has been shown in \textit{in vivo} experiments that five chaperones cooperate in a sequential ATP-dependent folding pathway [229]. Both \textit{in vitro} and \textit{in vivo} studies of protein folding have already revealed quite some details about their action [61, 62, 230, 231, 267] and other proteins assisting in protein folding like peptidyl-prolyl cis-trans isomerase and protein disulfide isomerase [232, 234]. Such studies may finally lead to the fine-tuning of the cellular machinery of protein folding. This would certainly lead to a considerable improvement for the manufacturing of correctly folded recombinant proteins.

\textbf{Cleavage of Fusion Proteins}

Fusion proteins exhibiting the desired biological activity may be used as such, if there are no special regulations requiring them to be transformed into their natural structure, as it is still common practice for proteins of therapeutical use. In the latter case the cleavage of the fusion complement has to be accomplished. For this purpose, the amino acid sequence of the fusion protein has to be constructed with a specific cleavage side.

Fusion proteins may be cleaved by chemical or enzymatic means. A chemical cleavage may be relatively unexpensive, but the rather drastic reaction conditions may lead to non-specific cleavage and denaturation of the target protein. Enzymatic methods are generally preferred, because specific cleavage sites may be introduced between the target protein and the fusion complement [63, 235]. Over 200 protein-cleaving enzymes are listed in a book of B. Keil [269]. Costs for the cleavage procedure may be reduced considerably by the development of an adequate standard cleavage sequence for a specific proteolytic enzyme. Table 1 contains an overview of available chemical and enzymatic cleavage methods and their respective cleavage sequences.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Method & Cleavage Sequence \hline
Chemical cleavage & BNPS-skatol, CAT, HIV, ompA, Kex 2 \hline
Enzymatic cleavage & Asp-N, Arg-C, Glu-C, Strep-Tags \hline
\end{tabular}
\end{table}

\textbf{Abbreviations used in Table 1:}

- BNPS-skatol = 3-bromo-3-methyl-2-(2-nitrophenyl mercapto)-3H-indol,
- CAT = chloroamphenicol acyltransferase, HIV = human immunodeficiency virus,
- ompA = outer membrane protein A, Kex 2 = yeast endoprotease
Table 1: Cleavage of fusion proteins

<table>
<thead>
<tr>
<th>Chemical methods</th>
<th>Cleavage agent</th>
<th>Cleavage-site sequence</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hydroxylamine</td>
<td>Asn ↓ Gly</td>
<td>[64]</td>
</tr>
<tr>
<td></td>
<td>Formic acid</td>
<td>Asp ↓ Pro</td>
<td>[65]</td>
</tr>
<tr>
<td></td>
<td>Acetic acid</td>
<td>Asp ↓ Pro</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>Cyanogen bromide</td>
<td>Met ↓</td>
<td>[67]</td>
</tr>
<tr>
<td></td>
<td>BNPS-skatol</td>
<td>Trp ↓</td>
<td>[68]</td>
</tr>
<tr>
<td></td>
<td>2-Iodo-benzoic acid</td>
<td>Trp ↓</td>
<td>[69]</td>
</tr>
<tr>
<td></td>
<td>N-Chlorosuccinimide</td>
<td>Trp ↓</td>
<td>[70]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enzymatic methods</th>
<th>Enzyme</th>
<th>Cleavage-site sequence</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carboxypeptidase A</td>
<td>Poly His ↓</td>
<td>[71]</td>
</tr>
<tr>
<td></td>
<td>Carboxypeptidase B</td>
<td>Poly Arg ↓, Poly Lys ↓</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td>Chymotrypsin</td>
<td>Trp ↓, Tyr ↓, Phe ↓</td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td>Collagenase</td>
<td>Pro-X ↓ Gly-Pro</td>
<td>[74]</td>
</tr>
<tr>
<td></td>
<td>Dipeptidyl aminopeptidase</td>
<td>X-Tyr ↓ (X not Pro)</td>
<td>[75]</td>
</tr>
<tr>
<td></td>
<td>Endoproteinase</td>
<td>Lys ↓</td>
<td>[76]</td>
</tr>
<tr>
<td></td>
<td>Enterokinase</td>
<td>Asp-Asp-Asp-Lys ↓</td>
<td>[77]</td>
</tr>
<tr>
<td></td>
<td>Factor Xα</td>
<td>Ile-Glu-Gly-Arg ↓</td>
<td>[78]</td>
</tr>
<tr>
<td></td>
<td>HIV-1 protease</td>
<td>CAT- ↓ -HIV-1 protease fusion, selfsplitting</td>
<td>[79, 80]</td>
</tr>
<tr>
<td></td>
<td>IgA-protease</td>
<td>Y-Pro ↓ X-Pro; Y = Pro, Ala, Gly, Thr X = Thr, Ser, Ala</td>
<td>[275]</td>
</tr>
<tr>
<td></td>
<td>Kallikrein</td>
<td>Pro-Phe-Arg ↓</td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td>Kex 2</td>
<td>-Lys ↓ Arg</td>
<td>[278]</td>
</tr>
<tr>
<td></td>
<td>ompA-Signal-peptidase</td>
<td>ompA-signal sequence ↓</td>
<td>[82]</td>
</tr>
<tr>
<td></td>
<td>Protein C</td>
<td>Phe-Thr-Phe-Arg, Leu-Ser-Thr-Arg, Pro-Glu-Leu-Arg</td>
<td>[268]</td>
</tr>
<tr>
<td></td>
<td>Renin</td>
<td>Tyr-Ile-His-Pro-Phe-His-Leu ↓ Leu</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td>S. aureus strain V8 protease</td>
<td>Glu ↓</td>
<td>[236]</td>
</tr>
<tr>
<td></td>
<td>Subtilisin</td>
<td>Ala-Ala-His-Tyr ↓</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td>Thrombin</td>
<td>Arg-Gly-Pro-Arg ↓</td>
<td>[85]</td>
</tr>
<tr>
<td></td>
<td>Trypsin</td>
<td>Arg ↓, Lys ↓</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>Ubiquitin peptidase</td>
<td>Ubiquitin ↓ Relaxin α-chain</td>
<td>[87]</td>
</tr>
</tbody>
</table>
The localization of proteins is obviously of major importance for downstream processing. If the protein is to be found in the cytoplasm, the concentrated cell mass has at first to be broken. The target protein, in consequence, has to be separated from a complex mixture of proteins, nucleic acids; eventually cellular compartments and debris. A more gentle cell disruption may be applied, if the target protein is secreted into the periplasm of Gram-negative bacteria. Obviously disruption is superfluous in the case of extracellular proteins.

**Separation of Cells and Cell Disruption**

Biomass is most often harvested by centrifugation or cross-flow filtration [6]. For *E. coli*, a genetic engineering approach has been developed for changing the properties of the cells in order to facilitate biomass recovery. A gene has been cloned which codes for a protein localized on the outer membrane surface and which is responsible for the flocculation properties. This modification resulted in a higher sedimentation velocity of the cells [88]. Similar interventions should also yield better properties for centrifugation and filtration.

Mechanical as well as non-mechanical techniques are applied for cell disruption on a large scale [6, 89]. The mechanical methods comprise high pressure homogenization and bead milling. Both processes may exhibit considerable protein losses due to incomplete protein liberation and thermal denaturation. Non-mechanical methods englobe chemical as well as biochemical processes. Cells may be permeabilized by organic solvents or by enzymatic lysis of the cell wall [90]. The use of organic solvents implies appropriate safety precautions [91], whereas enzymic processes are expensive on a large scale [90].

By means of genetic modifications it becomes possible to control cell lysis. The product of the *kil*-gene of the plasmid ColE1 may yield complete lyses of the cells [92]. A recent patent describes how the *kil*-gene, under the control of the *lac*-promoter may lyse the cells after induction with IPTG [93]. This strategy yields an expression system for which cell lyses may be induced under controlled conditions after the recombinant product has accumulated.

Based on the same principle, the lysis gene E of phage ΦX174 may be employed - under control of the *λ*-PL-promoter [94]. In the presence of the temperature-sensitive *λ*-repressor cl897, cell lysis may be induced by increasing the temperature to 42 °C.

**Secretion**

Protein export may circumvent some problems inherent to the expression of recombinant proteins [95]. It has already been discussed that the formation of inclusion bodies and protein degradation due to cytoplasmic proteases may be prevented. In addition, some
proteins may be lethal for a host when overproduced. Secretion obviously may reduce this phenomenon [82]. However, secretion also may yield incorrectly folded proteins especially in the case of eukaryotic proteins synthesized by bacteria.

Nevertheless, protein secretion into the medium considerably simplifies downstream processing. In this case the cells may be separated from the medium by e.g. centrifugation and the proteins may be isolated from the supernatant. The early separation of unbroken cells leads to reduced contamination by other proteins or cellular constituents - an obvious advantage. A minor disadvantage of this strategy is the need for handling large volumes of liquids.

The transport of proteins into the periplasm of Gram-negative bacteria represents a special case of protein secretion. After separation of biomass the product may be liberated from the periplasm prior to being isolated out of a relatively small volume [265]. Contamination due to cytoplasmic constituents may be avoided by gently removing the cell capsule and the outer membrane. This strategy has actually gained much attention. The periplasmic space representing about 20 to 40% of the cellular volume in the case of E. coli is large enough for the accumulation of large amounts of proteins. Besides the advantage that proteins in the periplasm are protected against the attack by cytoplasmic proteases (but not against outermembrane bound proteases [244], the environment of the periplasm favours the correct folding of proteins.

The gene of the target protein has to be coupled with an appropriate signal leader sequence and the host organism has to be provided with a cellular transport system in order to allow secretion of proteins. In addition, the target protein should not show properties preventing secretion.

**Signal Sequences:** Signal sequences commonly are short N-terminal peptides which enable proteins to use the sorting and transport systems of a particular organism. In bacteria signal peptides with a length of 15 to 30 amino acids are found [96,97]. These structures represent positively charged leader sequences allowing proteins to cross membranes. It may be mentioned that hemolysin of E. coli carries its signal peptide at the C-terminus. Between signal peptide and core protein a cleavage site is found at which the signal peptide is cleaved off by means of a specific signal peptidase after transport of the core protein has occurred.

Secreting proteins may show additional domains besides the signal sequence which may be necessary for a successful transport. In this category inner sequences are found which facilitate or even stop transport, like in the case of membrane proteins, giving rise to the development of an export competent conformation. Therefore, fusion with a signal peptide
### Table 2: Signal leader sequences for protein secretion

<table>
<thead>
<tr>
<th>Signal sequence</th>
<th>Target protein</th>
<th>M/P</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Host: <em>Escherichia coli</em></strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amy</td>
<td>Amylase, <em>B. stearothermophilus</em></td>
<td>M</td>
<td>[99]</td>
</tr>
<tr>
<td>bla</td>
<td>Proinsulin, human</td>
<td>-</td>
<td>[100]</td>
</tr>
<tr>
<td></td>
<td>IgG of mouse</td>
<td>-</td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td>β-Lactamase</td>
<td>M</td>
<td>[102]</td>
</tr>
<tr>
<td></td>
<td>Epidermal growth factor, rat</td>
<td>M/P</td>
<td>[103]</td>
</tr>
<tr>
<td></td>
<td>Triosephosphatase, chicken</td>
<td>M</td>
<td>[104]</td>
</tr>
<tr>
<td>cgt</td>
<td>alk. Phosphatase, α-amy lase</td>
<td>M</td>
<td>[282]</td>
</tr>
<tr>
<td>lamB</td>
<td>CD4 receptor of HIV</td>
<td>OM</td>
<td>[261]</td>
</tr>
<tr>
<td>malE</td>
<td>Gene 5 protein, phage M13</td>
<td>P</td>
<td>[105]</td>
</tr>
<tr>
<td></td>
<td>Klenow-polymerase</td>
<td>P</td>
<td>[106]</td>
</tr>
<tr>
<td></td>
<td>Nuclease A, <em>S. aureus</em></td>
<td>P</td>
<td>[106]</td>
</tr>
<tr>
<td>ompA</td>
<td>Colony stimulation factor, human</td>
<td>-</td>
<td>[107]</td>
</tr>
<tr>
<td></td>
<td>Superoxide dismutase, human</td>
<td>P</td>
<td>[82]</td>
</tr>
<tr>
<td></td>
<td>Interferon α2</td>
<td>-</td>
<td>[108]</td>
</tr>
<tr>
<td></td>
<td>Antiviral protein, <em>Mirabilis</em></td>
<td>M</td>
<td>[109]</td>
</tr>
<tr>
<td></td>
<td>α-Sarcin</td>
<td>P</td>
<td>[110]</td>
</tr>
<tr>
<td></td>
<td>Prokallikrein, human</td>
<td>P</td>
<td>[111]</td>
</tr>
<tr>
<td></td>
<td>Nuclease A, <em>S. aureus</em></td>
<td>P</td>
<td>[82]</td>
</tr>
<tr>
<td>ompF</td>
<td>β-Endorphin</td>
<td>M</td>
<td>[19]</td>
</tr>
<tr>
<td>pelB</td>
<td>Antibody VH-domain, mouse</td>
<td>OM</td>
<td>[276,280]</td>
</tr>
<tr>
<td>phoA</td>
<td>Trypsin inhibitor, bovine</td>
<td>-</td>
<td>[112]</td>
</tr>
<tr>
<td></td>
<td>Epidermal growth factor, human</td>
<td>-</td>
<td>[113]</td>
</tr>
<tr>
<td></td>
<td>Fusion : β-galactosidase-alk. phosphatase</td>
<td>M/P</td>
<td>[98]</td>
</tr>
<tr>
<td></td>
<td>α-neo-Endorphin</td>
<td>P</td>
<td>[114]</td>
</tr>
<tr>
<td></td>
<td>Fusion : MBP-β-galactosidase</td>
<td>P</td>
<td>[115]</td>
</tr>
<tr>
<td></td>
<td>Ribonuklease T1</td>
<td>P</td>
<td>[116]</td>
</tr>
<tr>
<td>phoS</td>
<td>Growth hormone release factor, human</td>
<td>P</td>
<td>[117]</td>
</tr>
<tr>
<td>spA</td>
<td>Parathyroid hormone, human</td>
<td>M</td>
<td>[118]</td>
</tr>
<tr>
<td></td>
<td>Insulin-like growth factor, human</td>
<td>M</td>
<td>[119]</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>Ovalbumin</td>
<td>-</td>
<td>[120]</td>
</tr>
<tr>
<td>Pullulanase</td>
<td>β-Lactamase</td>
<td>M</td>
<td>[121]</td>
</tr>
</tbody>
</table>
### Table 2: Signal leader sequences for protein secretion (continued)

<table>
<thead>
<tr>
<th>Host: <em>Escherichia coli</em> (cont.)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PreproInsulin (rat)</td>
<td>Proinsulin, rat</td>
<td>-</td>
<td>[122]</td>
</tr>
<tr>
<td>Enterotoxin LTA</td>
<td>Epidermal growth factor, human</td>
<td>-</td>
<td>[123]</td>
</tr>
<tr>
<td>Synthetic</td>
<td>Interferon α2</td>
<td>P</td>
<td>[124]</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>Metalloprotease, with helper protein</td>
<td>M</td>
<td>[125]</td>
</tr>
<tr>
<td>BRP</td>
<td>Insulin-like growth factor, human</td>
<td>M</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>β-lactamase &amp; α-amylase</td>
<td>M</td>
<td>[283]</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>M</td>
<td>[126]</td>
</tr>
<tr>
<td>hly</td>
<td>Hemolysin</td>
<td>M</td>
<td>[127]</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas cholesterol esterase</td>
<td>M</td>
<td>[246]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host: <em>Bacillus subtilis</em></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>alk. Protease</td>
<td>alkaline Phosphatase</td>
<td>M</td>
<td>[128]</td>
</tr>
<tr>
<td><em>B. amyloliquefaciens</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amy, <em>B. amyloliquefaciens</em></td>
<td>Amylase, <em>B. licheniformis</em></td>
<td>M</td>
<td>[129]</td>
</tr>
<tr>
<td>amy, prepro-peptide</td>
<td>Amylase &amp; human growth hormone</td>
<td>M</td>
<td>[247]</td>
</tr>
<tr>
<td>bla, <em>E. coli</em></td>
<td>Amylase, <em>B. licheniformis</em></td>
<td>M</td>
<td>[129]</td>
</tr>
<tr>
<td>Prepro-neutral-protease</td>
<td>Growth hormone, human</td>
<td>M</td>
<td>[130]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host: <em>Saccharomyces cerevisiae</em></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast killer toxin</td>
<td>α-Amylase, mouse</td>
<td>M</td>
<td>[248]</td>
</tr>
<tr>
<td>Prepro-α-factor</td>
<td>Viral proteins, human papillomavirus</td>
<td>M</td>
<td>[131]</td>
</tr>
</tbody>
</table>

**Abbreviations used in Table 2:**
may not be sufficient for the secretion of the fusion protein [98]. Table 2 contains a list of some signal sequences together with the target proteins for which they have been used.

*E. coli* is of interest as a secreting host organism, though even proteins naturally occuring in *E. coli* are hardly secreted. Known signal sequences like *bla* (*β*-lactamase), *malE* (maltose binding protein), *ompA* (outer membrane protein A) and *phoA* (alkaline phosphatase) originate from proteins being transferred into the periplasm of *E. coli*. Fusion proteins carrying these sequences are accordingly exported into the periplasm. However, this may not be true in any case. Even secretion into the medium has been observed. Thus, the structure of the fusion protein determines to quite some extent if a successful secretion may be achieved. Recent studies show evidence for the presence of a signal recognition particle in *E. coli* in analogy to eukaryotes [237, 238]. The signal sequences *spa* (*Staphylococcus* protein A) and *malE* are of particular importance since both may be used at the same time as affinity tags for facilitating the separation of the fusion protein.

**Secretion Systems:** Different organisms are used for aiming at secretion of proteins. Systems of bacterial origin include e.g. *Bacillus subtilis* [239], *Staphylococcus aureus*, *Streptomyces lividans* and last but not least *E. coli* [96]. Important eukaryotic secretion systems comprise different species of yeast like *Saccharomyces cerevisiae* [240], *Hansenula polymorpha* [241], *Kluyveromyces lactis* [242], *Yarrowia lipolytica* [243], of fungi like *Aspergillus oryzae*, and cellular strains of animal and human origin.

Not only the signal sequences used and the structural properties of the target protein are important in order to achieve successful secretion, but also the transport system and the composition of the cell membrane and the cell wall of the host organism.

There is increasing evidence that protein folding and protein export are competing processes in procaryotic cells. Disulfide bonds formed in the cytoplasm can lead to secretion incompetence [273]. In the case of *E. coli* there exists the possibility to gain control over the expression of certain proteins involved in the pathway of protein secretion [239]. These proteins belonging to the family of the *sec*-gene products can be expressed in parallel with the recombinant proteins in order to achieve an improved transport [97]. Modifications at the level of the cell membranes of *E. coli* lead to the appearance of so called leaky-mutants. With such mutants excretion of periplasmic proteins like alkaline phosphatase [132], β-lactamase [133], but also eukaryotic proteins like pro-insulin from rat [134] into the medium has been achieved. Therefore, the choice of an appropriate microbial strain [244] or cell line [245] represents an important merible with respect to the export competence. Another possibility consists in amplifying the amount of porin found in the outer membrane of *E. coli* [135]. However, such mutations often show the disadvantage of disturbing cell growth. Another interesting variation of this principle is the cloning of the
bacteriocin release protein (BRP). Human growth hormone which accumulated in the periplasm of *E. coli* was able to be exported into the medium upon induction of BRP [19].

**MODIFICATION OF PROTEINS FOR IMPROVEMENT OF SEPARATION AND PURIFICATION**

The separation and purification of proteins may be improved either by altering or by amplifying certain physico-chemical properties of the target protein by means of site-directed mutagenesis or by introduction of an affinity tag by means of a fusion complement. Adding a peptide or a protein as an affinity tag to the N- or C-terminus seems to be particularly attractive, since the tertiary structure of the target protein has not to be known and the primary structure is kept unchanged. In the case of a site-directed exchange of amino acids it should be known that these mutations occur at the surface of the protein in order to have an effect on downstream processing. In addition, changing the primary structure may affect the biological activity of the product. In general, protein fusion is easier to be accomplished than site-directed mutagenesis.

Most examples from literature are therefore to be found under the topic of fusion proteins. The fusion complement should exhibit a strong affinity for the complementary ligand used in the main process of separation. The advantage of the fusion method lies in the fact that in principle one generally applicable downstream processing strategy may be developed for a multitude of different target proteins. The additional cost for genetic engineering may be easily cushioned by savings on process development. Fusion proteins synthesized for the improvement of downstream processing often consist, beside the target protein and the affinity carrying protein, still of a signal leader sequence which should cause secretion of the fusion product into the periplasm in the case of Gram-negative bacteria or into the medium. Additionally, a peptide representing a specific cleavage site may be introduced between the target protein and the affinity carrying protein, if the desired protein has to be obtained in its native state. The particular problems of secretion and cleavage of fusion proteins has already been discussed in previous chapters.

Such fusion proteins will in future certainly be more and more commercialized without prior cleavage, if the biological activity of the target protein is not affected and if the fusion complement has no adverse effect on the particular application. It should also be taken into consideration that fusion proteins, that means proteins with two or multiple functions, are of strong interest for innumerable applications in areas such as general analytics, diagnostics, pharmaceutics as well as in health care [e.g. 19, 136, 137]. These multifunctional proteins are, however, so special in composition - e.g. an antibody fragment linked with a reporter enzyme - that an additional fusion with an affinity carrying protein may be envisaged in order to facilitate its recovery and purification [249]. It should be added that the construction of a fusion protein consisting of a purification tag and a
multiple repeat of the target protein each time separated by a cleavage site has also been proposed [250].

**Protein-Ligand Interaction**

It is important for the success of downstream processing strategies that fusion complements exhibit a strong specific interaction with the appropriate ligands used in the main separation procedure. The aim should be to obtain the fusion protein with near absolute purity in a single separation step. Appropriate pairs of combination of fusion complements and ligands are gathered in Table 3. The fusion complements carrying the site of affinity interaction are commonly called affinity handle, purification tag or affinity tag.

**Table 3: Potential combinations of complementary interactions for improving downstream processing of proteins**

<table>
<thead>
<tr>
<th>Modification of the target protein or fusion complement (peptide or protein)</th>
<th>Interacting ligands on the part of the separation process (not restricted to peptides and proteins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antigen</td>
<td>Antibody or antibody-fragment</td>
</tr>
<tr>
<td>Antibody or antibody fragment</td>
<td>Antigen</td>
</tr>
<tr>
<td>Enzyme or enzyme binding site</td>
<td>Substrate, inhibitor or substrate analogue</td>
</tr>
<tr>
<td>Substrat, inhibitor or substrate analogue</td>
<td>Enzyme or enzyme binding site</td>
</tr>
<tr>
<td>Receptor or receptor fragment</td>
<td>Hormone</td>
</tr>
<tr>
<td>Hormone</td>
<td>Receptor or receptor fragment</td>
</tr>
<tr>
<td>Lectin or lectin fragment</td>
<td>(Poly-)saccharide</td>
</tr>
<tr>
<td>Binding protein</td>
<td>Binding ligand (e.g. DNA, RNA ...)</td>
</tr>
<tr>
<td>(Poly-)histidine</td>
<td>Metal-chelate ligand</td>
</tr>
<tr>
<td>(Poly-)lysine, (poly-)arginine (poly-)glutamate, (poly-)aspartate</td>
<td>Charge ligands, Ion exchangers</td>
</tr>
<tr>
<td>(Poly-)phenylalanine</td>
<td>Hydrophobic ligand</td>
</tr>
<tr>
<td>(Poly-)cysteine</td>
<td>Thiol group</td>
</tr>
</tbody>
</table>
Purification tags obviously have to be peptides or proteins, whereas ligands may be derived from any source biological or chemical. The term affinity interaction is used for the capacity of complex formation between antibody and antigen (immunoaffinity), between receptor and messenger as well as between enzyme and substrate or inhibitor. The complexation of enzyme with substrate analogues and of e.g. (poly-)histidine sequences with chelating ligands is often called pseudo-affinity interaction. However, the term affinity is nowadays used in a quite broad sense.

Affinity-based separation methods are state of the art in downstream processing of proteins [138]. However, only affinity chromatography has been developed so far that a multitude of ligands bound on solid phases are commercially available [139]. Very high specificities may be obtained with monoclonal antibodies raised against fusion complements. Since the production of antibodies is quite expensive, the application is restrained to the recovery of expensive proteins which have to be obtained in extremely high purity. In this case fusion proteins offer the possibility that antibodies raised against the fusion complement may be used for the recovery of quite different target proteins. Thus, the costs for development and production of the antibody may be settled against the benefit from several products. A new trend is the expression of immunologically active antibody fragments in bacteria [140]. This should considerably diminish the costs of production of specific antibody activities. This would also allow for a reversal of the commonly used method leading to the situation that the fusion complement may be the antibody and the ligand may act as the antigen. However, a problem commonly encountered for using the principle of immuno-affinity is that the antibody-antigen complexes only dissociate under quite drastic conditions.

Enzymes and their binding sites are quite convenient fusion complements, if these are known to form stable complexes with natural or synthetic substrates or inhibitors, respectively. A reversal of this principle is possible if a peptide may be identified as a potent inhibitor. All kinds of substrate affinity are quite important for downstream processing of enzymes. The most extensive experiences for the use of affinity interactions have been gathered in this area. Reactive dyes have found particular attention as chemically stable ligands [141]. These had been introduced mainly as coenzyme analogues into affinity separation techniques. Quite a number of this ligands bound to solid phases are commercially available. In principle, there is obviously an innumerable number of candidates on both sides, the purification tag and the ligand.

Highly specific interaction may be expected from the combination of a receptor and its respective messenger. If the messenger is a peptide hormone the principle may be reversed.

Saccharides are potent ligands binding to appropriate lectins. If the gene of the appropriate lectin is known, this interaction may be used in protein recovery as well.
The class of binding proteins shall comprise all those proteins which do not fit into the categories already mentioned. Examples for such proteins are those which exhibit pseudo-immuno-affinity interactions like the *Staphylococcus* proteins A [142] and G, which bind to specific gene sequences (DNA or RNA) as well as proteins which interact with certain chemical groups or molecules. New methods from molecular genetics like reverse transcription and the polymerase chain reaction allow for producing (short chain) RNA, which binds certain proteins or peptides with high specificity [143,144], by repeated cycles of mutation/selection/amplification. This technique might be used in order to synthesize specific (RNA-) ligands for any peptide which is to be used as a fusion complement.

An extremely interesting method for improving the separation and purification of proteins is based on the interaction of metal chelate ligands with, particularly, histidine side chains of proteins, because it is rather simple to be applied. Although this interaction is known since 1975, it has only found deeper interest because of the evolving genetic engineering methods. These methods permit either to introduce histidine residues at the surface of proteins by site-directed mutagenesis or to introduce a histidine-rich sequence at the terminus of the target protein. With respect to the specificity of the interaction it is quite helpful to know that histidine represents one of the most rare proteinogeneous amino acids. The most commonly used chelating ligand is the three-dentate iminodiacetic acid (IDA). Appropriate supports are easily derivatized with this chelating group [145] and respective chromatographic phases are offered by several companies [146]. Additional chelating ligands are N,N,N'-Tris (carboxymethyl) ethylenediamine (TED) [145], nitritotri-acetic acid (NTA) [147] - actually N-(5-amino-1-carboxypentyl) iminodiacetic acid, a lysine derivative - and ethylenediamine-N,N'-diacetic acid (EDDA) [148]. Metal ions which have most commonly been used comprise Cu(II), Ni(II), Zn(II) and Fe(III). Cu(II)-IDA groups form already stable complexes in the presence of only one accessible histidine side chain, whereas for Zn(II)-IDA two vicinal histidine residues are required [149]. However, recent results have shown that there are exceptions from this simple rule [150]. The Ni(II)-NTA group seems to be particularly suited to complexing proteins with several histidine residues in series [151]. The chelate-protein complexes are stable in the neutral pH-range. Complex dissociation may be forced by decreasing the pH or by means of adding ammonium ions, imidazole, histidine or other complexing agents which compete with the protein for the chelating group. The application of this principle in chromatography is known under the term immobilized metal affinity chromatography (IMAC) which is well documented [146, 149]. For modification of the target protein, peptides of several histidine residues are the obvious fusion complements to be used in order to achieve a high selectivity of separation.

Altering the properties of proteins by means of fusion with amino acid oligomers or with self-repeating amino acid sequences can be done quite easily by genetic engineering. Oligomers of amino acids with charged side chains (lysine, arginine as well as glutamic-
and aspartic acid) may be used in order to recover proteins selectively by means of oppositely charged groups like ion exchangers. Hydrophobic interaction may be used for the recovery of proteins derivatized with oligomers of hydrophobic amino acids. The reactive thiol group of cysteine is able to form reversible covalent bonds with ligands.

**Downstream Processing**

Separation processes and purification steps of particular interest for downstream processing of recombinant proteins shall be discussed in this chapter. In principle, this area is well documented [e.g. 8-15, 152], however, it evolves rapidly. With increasing amount of proteins of high purity, large scale processing [13, 153] and continuous operation [5] are gaining considerable interest.

A survey on processes applied for the separation and purification of recombinant proteins is given in Table 4. The largest choice of commercially available solid supported ligands can be found for processes based on adsorption. A rich chemistry for the activation of solid surfaces and polymer matrices is known in order to bind particular ligands in such a way that they may interact with proteins. Therefore, chromatography is the most commonly used technique for protein separation and purification. Unfortunately, the term chromatography is used nowadays for quite a number of different techniques.

**Table 4 : Processes for downstream processing of proteins**

<table>
<thead>
<tr>
<th></th>
<th>Processes for downstream processing of proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Solid-phase extraction</td>
</tr>
<tr>
<td>1.1</td>
<td>Adsorption</td>
</tr>
<tr>
<td>1.2</td>
<td>Chromatography</td>
</tr>
<tr>
<td>2</td>
<td>Liquid-phase extraction</td>
</tr>
<tr>
<td>2.1</td>
<td>Aqueous two-phase extraction</td>
</tr>
<tr>
<td>2.2</td>
<td>Microemulsion two-phase extraction</td>
</tr>
<tr>
<td>3</td>
<td>Precipitation</td>
</tr>
<tr>
<td>4</td>
<td>Membrane processes</td>
</tr>
<tr>
<td>4.1</td>
<td>Cross-flow filtration (micro- or ultrafiltration)</td>
</tr>
<tr>
<td>5</td>
<td>Electrophoretic processes</td>
</tr>
</tbody>
</table>
If proteins are modified by genetic engineering with respect to downstream processing, it is in most cases done by introducing affinity interactions with the aim that a stable complex of the protein and a carrier-fixed ligand may be obtained. The respective adequate processing step is commonly called chromatography, also called replacement chromatography, but represents a process of adsorption/desorption. By means of this technique, the capacity of the solid phase can be used to a maximum. The separation selectivity is optimized by the choice of different eluents applied in series. Elution is mainly carried out in fixed beds in order to achieve a concentrating effect as high as possible in addition to separation owing to the small degree of backmixing, whereas charging may happen as well in well-mixed apparatus [154] or in fluidized beds [155]. Well-mixed systems are particularly suited, if the target protein is to be recovered in the presence of biomass or cell debris [156-158]. Since the diffusional resistance of proteins is quite high, small macroporous particles have to be applied. However, the operation of beds of small particles may be limited by a high pressure drop. Therefore, columns with small aspect ratio (e.g. radial-flow columns) have gained considerable attention. In last consequence this leads to the application of functionalized porous membranes [159-161].

In the area of liquid phase extraction, aqueous two-phase systems have proved successful not only for the separation of biomass and cell debris but also for the separation of proteins [e.g.15, 162]. Since phase separation occurs due to the incompatibility among water-soluble polymers or of these with salts, affinity interactions can be used, if the respective ligands are bound to the respective polymers. A polymer of central importance is polyethylene glycol (PEG). This polymer may be easily derivatized with groups like coenzymes, reactive dyes, chelate- and other affinity ligands [163]. The derivatized PEG is able to drag the protein into the polymer-rich phase [164-168].

Another system for liquid phase extraction consists of an aqueous phase in contact with a microemulsion - the latter is also called reverse micellar phase. Microemulsions are thermodynamically stable systems consisting of water, surface active agents and unpolar solvents. Proteins may be solubilized in the so-called reverse micelles. Proteins may be incorporated preferentially into the microemulsion phase by synthesizing surfactants of which the hydrophilic group represents the affinity ligand [169].

The precipitation by means of ammonium sulfate or acetone is a classical process for concentrating and fractionating proteins. A precipitation of higher selectivity and requiring less auxiliary substances may be achieved, if bifunctional ligands or functionalized polymers are used as precipitating agents. Polymers like polyethyleneimine (PEI) or chitosan are well suited in order to precipitate negatively charged proteins. Such polymers may be coupled with affinity ligands like in the case of polyethylene glycol, and may be used for affinity precipitation [166, 170, 171].
Common membrane processes like ultrafiltration are unsuitable for the fractionation of proteins, because of a quite diffuse pore size distribution of membranes and the interaction of different proteins in the stagnant boundary layer [172]. Only if the molar mass of the target protein would be enlarged considerably by artificial means, it could be separated from other proteins present. This may be achieved most simply and selectively by using high molar mass polymers or microscopic supports derivatized with affinity ligands which are effectively retained by porous membranes. When the crude protein mixture is pumped into the ultrafiltration plant, the target protein forms a complex with the support. The complex as such is retained, whereas undesired proteins may leave the ultrafiltration unit. After a washing cycle, the target protein may be liberated by means of the addition of a competing complexing agent. For this purpose, supports have been used in the form of soluble polymers as well as of finely divided porous solid particles [166].

Since the charge and the isoelectric point, respectively, of a protein may be altered quite easily by genetic engineering methods, electrophoretic separation techniques may be of interest, too. For the purpose of production, continuously operating processes would be most appropriate [173]. Respective apparatus with uncompartimented cells are already available [174, 175]. Cells divided by ampholyte membranes and equipped with separate loops for liquid recycling may be used for producing large amounts of proteins discontinuously. Thus, minor differences with respect to the isoelectric point are sufficient in order to obtain proteins of high purity [176].

**Survey of Applications**

Examples with respect to the use of genetic engineering methods for the improvement of protein separation and purification are gathered in Table 5. Each new issue in the table is initiated by a hyphen. If the hyphen is omitted in one of the columns, the last entry above the actual position is still valid. Some of the signs used in Table 5 have to be explained. The sign "&" signifies that different experiments have been carried out, whereas the sign "+" in the column fusion complement is used when both N- and C-terminal fusions were applied to one target protein.

The entries in Table 5 are arranged according to the kind of interaction. It has to be mentioned that, when protein A (SPA) has been used as a fusion complement, this is to be found under the category of immunoaffinity because it binds to immunoglobulin G (IgG). However, it does not interact with the antigenic determinant, but with the constant Fc-region. This kind of binding is commonly called pseudo-immuno affinity interaction.

With a few exceptions, almost all of the proteins mentioned in Table 5 have been produced by expression in the domestic bug of the molecular geneticists, in *E. coli*. Therefore, the
potential for secretion is quite restricted. By fusion of the target proteins with secretory proteins or signal leader sequences, the respective fusion proteins may be directed into the periplasmic space of *E. coli*. For secretory purposes the genes of e.g. ompA, protein A, protein G or the maltose binding protein have been used for complementation. Eukaryotes and Gram-positive prokaryotes are better suited, if export into the medium is desired. However, protein release into the medium has been observed in rare cases with *E. coli*, too. Otherwise, Table 5 does not contain any indication about signal sequences used for inducing secretion nor about the introduction of special sequences serving as cleavage sites for the fission of fusion proteins. Nevertheless, the respective auxiliaries are indicated which have been used for cleaving the fusion protein - as far as this topic was a subject of investigation. Two cases are included (Table 5, No. 1.10, 2.13) the objective of which has been to obtain fusion proteins of enzymes for facilitating their immobilization. Since the fusion complement serves for binding with the support, the question in this case is similar to that where the objective is downstream processing. However, the very product is the fusion protein exhibiting the activity of the target enzyme.

The β-galactosidase of *E. coli* does not appear particularly noticeable in Table 5, because only recent literature (since ca. 1984) is reviewed. Respective fusion vectors with or without protease cleavage site are commercially available [177]. All fusions with β-galactosidase of *E. coli* are suited *a priori* for the improvement of downstream processing since its substrate affinity for e.g. APTG may be exploited. The compilation of the target proteins clearly shows the great interest for the expression of eukaryotic proteins and particularly those of human origin in rapidly growing prokaryotes. Several of the target proteins mentioned like β-lactamase, SPA-ZZ, β-galactosidase and galactokinase are simply model proteins which have been taken for testing purposes. Table 5 only contains two examples (Table 5, No. 4.11, 7.1) in case of which site-directed mutagenesis has been applied for improving downstream processing. In all other cases purification tags have been introduced by means of fusion.

The ideal fusion complement serving as affinity handle should be able to form a stable complex with an inexpensive, chemically stable and easily derivatized ligand with an absolute specificity. However, the complex should dissociate under mild conditions. In addition, it is desirable that both components affinity handle and ligand would be of simple structure. With respect to these requirements, the FLAG™-peptide may be mentioned. It consists of a sequence of only 8 amino acids (AspTyrLysAsp₄Lys-) and may be used for N-terminal fusion. The first four amino acids represent the antigen with the help of which a fusion protein may bind to the monoclonal antibody (mouse anti-FLAG IgG M1) as the ligand. The following sequence (Asp₃Lys-) represents the specific binding site for enterokinase which can be used for splitting off the octapeptide from the target protein. Although this case deals with an antibody-antigen complex, dissociation may be achieved under mild conditions because the interaction depends on the presence of calcium ions.
**Table 5:** Examples for the improvement of downstream processing of recombinant proteins

<table>
<thead>
<tr>
<th>Interaction No.</th>
<th>Example for ligand (purification tag)</th>
<th>Target protein</th>
<th>Host organism</th>
<th>Secr.</th>
<th>Cleavage</th>
<th>Immuno affinity</th>
<th>Cleavage References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>- FLAG™-peptide</td>
<td>anti-FLAG mAb</td>
<td>E. coli</td>
<td>+</td>
<td>-</td>
<td>Enterokinase</td>
<td>[189]</td>
</tr>
<tr>
<td>1.1.1</td>
<td>- scFv-PhoA</td>
<td>anti-FLAG mAb</td>
<td>S. cerevisiae</td>
<td>ex</td>
<td>-</td>
<td>Enterokinase</td>
<td>[249]</td>
</tr>
<tr>
<td>1.2</td>
<td>- TNFα poresist</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>(+)</td>
<td>-</td>
<td>Enterokinase</td>
<td>[251]</td>
</tr>
<tr>
<td>1.3</td>
<td>- α-defensin</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>(+)</td>
<td>-</td>
<td>Enterokinase</td>
<td>[190]</td>
</tr>
<tr>
<td>1.4</td>
<td>- α-defensin</td>
<td>anti-β-Gal mAb</td>
<td>S. cerevisiae</td>
<td>MP</td>
<td>+</td>
<td>Collagenase</td>
<td>[191]</td>
</tr>
<tr>
<td>1.5</td>
<td>- TrpLE</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>MP</td>
<td>+</td>
<td>Collagenase</td>
<td>[192, 182]</td>
</tr>
<tr>
<td>1.6</td>
<td>- Lipase</td>
<td>anti-β-Gal mAb</td>
<td>S. aureus</td>
<td>ex</td>
<td>+</td>
<td>Collagenase</td>
<td>[193]</td>
</tr>
<tr>
<td>1.7</td>
<td>- α-defensin</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>MP</td>
<td>+</td>
<td>Collagenase</td>
<td>[194]</td>
</tr>
<tr>
<td>1.7.1</td>
<td>- SPA-Gal</td>
<td>anti-β-Gal pAb</td>
<td>E. coli</td>
<td>MP</td>
<td>+</td>
<td>Collagenase</td>
<td>[195]</td>
</tr>
<tr>
<td>1.8</td>
<td>- SPA-EE</td>
<td>anti-β-Gal mAb</td>
<td>S. aureus</td>
<td>ex</td>
<td>+</td>
<td>Collagenase</td>
<td>[64, 178]</td>
</tr>
<tr>
<td>1.9</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>S. aureus</td>
<td>ex</td>
<td>+</td>
<td>Collagenase</td>
<td>[43]</td>
</tr>
<tr>
<td>1.10</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>ex</td>
<td>+</td>
<td>Collagenase</td>
<td>[191]</td>
</tr>
<tr>
<td>1.11</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>S. aureus</td>
<td>ex</td>
<td>+</td>
<td>Collagenase</td>
<td>[196]</td>
</tr>
<tr>
<td>1.11.1</td>
<td>- SPA-IgG</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>+</td>
<td>ex</td>
<td>SPA-IgG-BD</td>
<td>[86]</td>
</tr>
<tr>
<td>1.12</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>+</td>
<td>ex</td>
<td>SPA-IgG-BD</td>
<td>[71, 195]</td>
</tr>
<tr>
<td>1.13</td>
<td>- SPA-EE</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>+</td>
<td>ex</td>
<td>SPA-IgG-BD</td>
<td>[74]</td>
</tr>
<tr>
<td>1.14</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>E. coli</td>
<td>+</td>
<td>ex</td>
<td>SPA-IgG-BD</td>
<td>[74]</td>
</tr>
<tr>
<td>1.15</td>
<td>- SPA-ZZ</td>
<td>anti-β-Gal mAb</td>
<td>S. aureus</td>
<td>+</td>
<td>ex</td>
<td>SPA-IgG-BD</td>
<td>[64, 178]</td>
</tr>
</tbody>
</table>

**References:**
- [189]
- [249]
- [251]
- [190]
- [191]
- [192, 182]
- [193]
- [194]
- [195]
- [64, 178]
- [43]
- [191]
- [196]
- [86]
- [71, 195]
- [74]
- [64, 178]
<table>
<thead>
<tr>
<th>Substrate affinity</th>
<th>Enzyme or enzyme binding site</th>
<th>Substrate, inhibitor or substrate analogue</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>- β-Galactosidase</td>
<td>- APTG - DNA-BP(ss) &amp; R6K-RI</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>- ABTG</td>
<td>- TRP</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>- CAT</td>
<td>- HDNO</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>- Chloramphenicol</td>
<td>- Calcitonin etc.</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>- ANF</td>
<td>- E. coli</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>- GST</td>
<td>- SPA or SPG</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>- Glutathion</td>
<td>- P. falciparum Antigens</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>- PTP</td>
<td>- E. coli</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>- CFTR</td>
<td>- E. coli</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>- ChoBD</td>
<td>- aFGF</td>
<td></td>
</tr>
<tr>
<td>2.11</td>
<td>- DEAE</td>
<td>- E. coli</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>- Glucose Amylase SBD</td>
<td>- Starch</td>
<td></td>
</tr>
<tr>
<td>2.13</td>
<td>- Exoglucanase CBD</td>
<td>- Glucosidase (partial)</td>
<td></td>
</tr>
<tr>
<td>2.14</td>
<td>- Cyclodextrin glycosyl-transferase</td>
<td>- α-Cyclodextrin-agarose</td>
<td></td>
</tr>
</tbody>
</table>

Staph. protease

Thrombin etc.

Factor X_a

Enyzimm

[197-199]

[191]

[200]

[201]

[202]

[203]

[204]

[224]

[256]

[257]

[205]

[206-208]

[282]
<table>
<thead>
<tr>
<th>General binding affinity</th>
<th>3.1</th>
<th>β-Galactosidase</th>
<th>PEG 4000</th>
<th>SPA or SPG</th>
<th>E. coli</th>
<th>-</th>
<th>[181, 182]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptor, lectin, binding protein</td>
<td>3.2</td>
<td>(AlaTrpTrpPro)1-3</td>
<td>PEG 4000</td>
<td>SPA-ZZ</td>
<td>E. coli</td>
<td>-</td>
<td>[183, 255]</td>
</tr>
<tr>
<td>Binding ligand</td>
<td>3.3</td>
<td>Phosphatase-BP</td>
<td>Hydroxylapatit</td>
<td>hGH RF</td>
<td>E. coli</td>
<td>+</td>
<td>[117]</td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td>Streptavidin</td>
<td>Biotin</td>
<td>α-Antitrypsin</td>
<td>E. coli</td>
<td>ex</td>
<td>[209]</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>SPG &amp; SPG-fragments</td>
<td>hSA</td>
<td>-</td>
<td>E. coli</td>
<td>+</td>
<td>[210]</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
<td>Maltose-EP</td>
<td>Starch</td>
<td>β-Gal &amp; paramyosin</td>
<td>E. coli</td>
<td>+</td>
<td>Factor Xa</td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>(AlaTrpTrpPro)c.s</td>
<td>PEG 4000</td>
<td>SPA-R</td>
<td>E. coli</td>
<td>-</td>
<td>[123]</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>Phosphate-BP</td>
<td>Hydroxylapatit</td>
<td>hGH RF</td>
<td>E. coli</td>
<td>+</td>
<td>[1115]</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>Streptavidin</td>
<td>Biotin</td>
<td>α-Antitrypsin</td>
<td>E. coli</td>
<td>ex</td>
<td>[210]</td>
</tr>
<tr>
<td></td>
<td>3.10</td>
<td>SPG &amp; SPG-fragments</td>
<td>hSA</td>
<td>-</td>
<td>E. coli</td>
<td>+</td>
<td>[210]</td>
</tr>
<tr>
<td>Metal-chelate affinity</td>
<td>4.1</td>
<td>His\textsubscript{6}</td>
<td>Ni(II)-NTA</td>
<td>DHFR</td>
<td>E. coli</td>
<td>-</td>
<td>CPA</td>
</tr>
<tr>
<td></td>
<td>4.2</td>
<td>His\textsubscript{6}</td>
<td>Ni(II)-IDAN</td>
<td>HIV-1 RT</td>
<td>E. coli</td>
<td>+</td>
<td>HIV-1 protease</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>His\textsubscript{6}</td>
<td>Ni(II)-NTA</td>
<td>HIV-1 RT &amp; -protease</td>
<td>E. coli</td>
<td>+</td>
<td>HIV-1 protease</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>His\textsubscript{6}</td>
<td>Ni(II)-NTA</td>
<td>HIV-1 RT &amp; -protease</td>
<td>E. coli</td>
<td>+</td>
<td>HIV-1 protease</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>His\textsubscript{6}</td>
<td>Ni(II)-NTA</td>
<td>HIV-1 RT &amp; -protease</td>
<td>E. coli</td>
<td>+</td>
<td>HIV-1 protease</td>
</tr>
<tr>
<td>Metal-chelate</td>
<td>4.6</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>4.7</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>4.8</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>4.9</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>4.10</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>4.11</td>
<td>His\textsubscript{6}</td>
<td>Zn(II)-NTA &amp; Ni(II)-NTA</td>
<td>F\textsubscript{v} fragment (mouse-IgA)</td>
<td>E. coli</td>
<td>+</td>
<td>[47]</td>
</tr>
</tbody>
</table>

**Improving Downstream Processing**
<table>
<thead>
<tr>
<th>Charge ligands</th>
<th>Kation exchanger</th>
<th>β-Urogastron</th>
<th>E. coli</th>
<th>CP B</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Poly-)Arg,Glu,Asp</td>
<td>Arg5</td>
<td>-</td>
<td>-</td>
<td>[72, 217, 218]</td>
</tr>
<tr>
<td>5.1</td>
<td>Asp5-16</td>
<td>PEI</td>
<td>-Galactosidase</td>
<td>E. coli</td>
</tr>
<tr>
<td>5.2</td>
<td>Glu5</td>
<td>-</td>
<td>hGH</td>
<td>-</td>
</tr>
<tr>
<td>5.3</td>
<td>-</td>
<td>-</td>
<td>E. coli</td>
<td>-</td>
</tr>
<tr>
<td>(Poly-)phenylalanine</td>
<td>Phenyl-Group</td>
<td>-Galactosidase</td>
<td>E. coli</td>
<td>-</td>
</tr>
<tr>
<td>Hydrophobic ligand</td>
<td>6.1</td>
<td>Phe11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6.1</td>
<td>Cys (Ser24Cys), SDM</td>
<td>Thiol-Group</td>
<td>His64Ala Subtilisin</td>
<td>B. subtilis</td>
</tr>
<tr>
<td>7.1</td>
<td>Cys4</td>
<td>Galactokinase</td>
<td>-</td>
<td>ex</td>
</tr>
<tr>
<td>7.2</td>
<td>-</td>
<td>-</td>
<td>E. coli</td>
<td>-</td>
</tr>
<tr>
<td>(Poly-)cysteine</td>
<td>ThioI-Group</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Abbreviations:

- AB(P)TG = p-aminobenzyl(phenyl)-β-D-thiogalactoside
- ANF = atrial natriuretic factor
- BD = binding domain
- BP = binding protein
- BPTI = trypsin inhibitor from bovine pancreas
- CAT = chloramphenicol acetyl transferase
- CBD = cellulose binding domain
- CFTR = cystic fibrosis transmembrane conductance regulator
- ChoBD = cholin binding domain of S. pneumoniae
- N-acetyl-
- L-alanine amidase (autolysin)
- CP = carboxypeptidase
- CRD = carbohydrate-recognition domain
- DHFR = mouse dihydrofolate reductase
- DPAP = dipeptidyl aminopeptidase
- EN = endonuclease
- EnzImm = fusion aiming at immobilization of the target enzyme
- ex = export into the medium
- αFGF = acidic fibroblast growth factor
- GH = growth hormone
- GHRF = growth hormone release factor
- GM-CSF = colony stimulating factor for granulocytes and macrophages
- GST = glutathion-S transferase
- G5P = gene-5 protein, binds single-strand DNA
- h.. = human
- HDNO = 6-hydroxy-D-nicotin oxidase
- IB = inclusion bodies
- IDA = iminodiacetic acid
- IGF = insulin-like growth factor
- IgG = immunoglobulin G
- IL-2 = Interleukin 2
- LMM = C-terminal fragment of light meromyosin from rabbit
- mAb = monoclonal antibody
- MP = membrane protein
- NTA = Nitrilotriacetic acid
- pAB = polyclonal antibody
- PAP = placenta alkaline phosphatase
- PEG = Polyethylene glycol
- PhoS = phosphate binding protein
- RT = reverse transcriptase
- R6K-RI = RK6 replication initiator
- PRC = photosynthetic reaction centre (from R. sphaeroides)
- PTP = protein tyrosine phosphatase from rat brain
- SA = serum albumin
- SBD = starch binding domain
- scFv = single-chain antigen binding protein
- variable mAb-domain for recognition of human erbB-2 protein
- SDM = site-directed mutagenesis
- SNA = Staphylococcus aureus nuclease A
- SPA = Staphylococcus protein A
- SPA-EE(ZZ) = dimer of a (synthetic) IgG binding fragment of SPA
- SPG = Staphylococcus protein G
- SPG-B1B2 = fusion of two hSA binding domains from SPG
- TNFα = tumor necrosis factor alpha
- TrpLE = leader sequence of the trp-operon
- USF = human gene-specific transcription factor
- Z = synthetic IgG binding domain of SPA
The fusion protein is only bound in the presence of calcium ions, whereas the complex dissociates when calcium ions are withdrawn. Since the FLAG-peptide is small and very hydrophilic, it can be expected that it may be found at the protein surface and thus may be accessible [178]. The dependence on the availability of specific antibodies, however, has to be taken into account for an economic analysis of the separation process.

Fusion complements based on protein A from Staphylococcus aureus are of interest in many respects. This protein is stable against proteolysis and forms stable complexes with the constant region (F_\text{C}) of IgG of different mammals. The wild type gene consists of a signal leader sequence for secretion, five IgG-binding domains (E,D,A,B,C) and an additional binding domain for the integration of protein A in the cell wall of S. aureus. Fusion complements may be selectively assembled by taking e.g. the signal sequence together with a single or a repeated sequence of binding domains [142]. The complement SPA-ZZ is such a designed complement consisting of the signal leader sequence and two B-domains containing two point mutations. These point mutations (Asn\text{Gly}→Asn\text{Ala}) were introduced in order to protect the B-domain against hydrolysis by hydroxylamine. A site directed mutation at the N-terminus of e.g. IGF-I (Met→Asn) is sufficient for the introduction of a cleavage side (Asn\text{Gly}) to allow the fusion protein to be split by means of hydroxylamin. By using both fusion complements SPA-EE and -ZZ it has even been observed that about 80% of the fusion protein with IGF-I or -II expressed in E. coli were exported into the medium [64, 119]. Fermentation capacities up to 1000 L have already been applied for the production of IGF-I [179]. Vectors for the assembly of SPA-fusions are commercially available [177]. Plasmids with a temperature - inducible promoter have been described [252] as well as one which permits a C-terminal fusion with SPA-ZZ [253].

The periplasmic maltose binding protein of E. coli is another commonly used fusion complement. It may be used for inducing the secretion into the periplasm. The fusion product adsorbs on cross linked starch, whereas elution may be achieved in the presence of maltose. A disadvantage of this purification tag is its relatively high molar mass (about 41 kDa) in comparison with e.g. SPA-ZZ (about 14 kDa) or the FLAG-peptide (< 1 kDa). However, an advantage which should not be underestimated is given by the fact that an inexpensive ligand can be used.

An inexpensive ligand is also used in the case of the metal-chelate technique. The complementation of the target protein with histidine-rich sequences can easily be achieved. With the introduction of vicinal histidine residues and their interaction with Ni(II)- or Zn(II)-activated chelating groups, a high selectivity may be obtained [254]. Some cloning kits are commercially available which are based on a His_\text{6} fusion complement [180].

A categorization according to the kind of separation or purification process is not given in Table 5. This is due to the fact that other processes than adsorption and chromatography,
respectively, have rarely been applied. Extraction in aqueous two-phase systems may successfully be used, if the fusion complement interacts specifically with the soluble polymer of the phase system (e.g. PEG). It is known that β-galactosidase of *E. coli* is characterized by an exceptionally high distribution coefficient in favour of the PEG-rich phase of PEG-salt systems. If e.g. protein A or G are linked with β-galactosidase, an essentially more favourable distribution may be found compared with the extraction of the original proteins [181, 182]. Tryptophan side chains on the surface of β-galactosidase are thought to be responsible for the favourable partitioning of this enzyme. This assumption has impressively been proved by the design of a fusion complement consisting of the peptide AlaTrpTrpPro ("partitioning peptide") and the use of oligomers of this sequence [183, 255]. In such cases specific affinity ligands are not required. However, the interaction with specific affinity ligands bound to a polymer of the aqueous two-phase system is of more general application. The influence of X→His point mutations on the affinity partitioning in Co(II)-IDA-PEG-doted PEG-salt systems has been studied. If histidine residues are introduced into the protein backbone in such a way that two of them may be found in vicinal position on an α-helical structure (His-X₃-His) on the protein surface, a considerable improvement of the distribution in favour of the polymer-rich phase can be expected [184].

Only two groups of authors report about the application of affinity precipitation. Fusion proteins of β-galactosidase of *E. coli* with polyaspartate sequences of various length have been precipitated by means of polyethylenimine [185, 186]. Relatively long polyaspartate sequences were required in order to achieve an appreciable accumulation of precipitate - after nucleic acids had been broken down [185]. Ligands with more selective interaction would certainly be better suited as precipitating agents [e.g. 187]. A fusion protein of human P-glycoprotein and β-galactosidase has been recovered by immunoprecipitation [74]. In this case the fusion protein was complexed with polyclonal antibodies against β-galactosidase of rabbit and was precipitated with protein A bound to sepharose which means that the fusion protein was recovered adsorbed to the sepharose particles. Auto-precipitation may be induced by utilizing the C-terminal fragment of light meromyosin (LMM) from rabbit fast skeletal muscle as the fusion complement. The myosin heavy chains undergo coiled-coil interaction. Such fusion proteins precipitate in dilute salt solutions whilst they are soluble in concentrated salt solutions [258].

If the fusion protein is cleaved in order to obtain the target protein in its native form, the fusion complement may be removed by means of the same strategy used for the separation of the fusion protein [18]. This principle confers a certain charm to the strategy of fusion. One problem remains to be solved, that is the removal of the protease used for cleaving the fusion product. The immobilization of proteases on macroporous supports of small particle size might eventually solve this problem in an economic manner [188, 235].
This short review under the subject of genetic engineering methods for improvement of downstream processing of proteins has shown that extremely interesting strategies have been developed which are equally being applied in both areas research and production. However, it has to be mentioned that still a lot of problems have to be solved. The practical application of these strategies still holds many surprises due to superficial knowledge about protein expression, folding, sorting, and post-translational processing. The safe knowledge accumulated so far together with the rapid progression in this area, however, will certainly lead to a broad application of the strategies briefly outlined here for the production of recombinant proteins.

REFERENCES

22. Nordström, K., Uhlin, B.E.: Runaway-replication plasmids as tools to produce large quantities of proteins from cloned genes in bacteria, Bio/Technol. 10 (1992) 661-666


57. Dill, K.A.: Dominant Forces in Protein Folding, Biochem. 29 (1990) 7133-7155


70. Shechter, Y., Patchornik, A., Burstein, Y.: Selective Chemical Cleavage of Tryptophanyl Peptide Bonds by Oxidative Chlorination with N-Chlorosuccinimide, Biochem. 15 (1976) 5071-5075


115. di Guan, C., Li, P., Riggs, P.D., Inouye, H.: Vectors that Facilitate the Expression and Purification of Foreign Peptides in *Escherichia coli* by Fusion to Maltose-Binding Protein, Gene **67** (1988) 21-30

116. Fujimura, T., Tanaka, T., Kanako, O., Morioka, H., Uesugi, S., Ikehara, M., Nishikawa S.: Secretion of recombinant ribonuclease T₁ into the periplasmic space of *Escherichia coli* with the aid of the signal peptide of alkaline phosphatase, FEBS **269** (1,2) (1990) 71-74


120. Fraser, T.H., Bruce, T.J.: Chicken ovalbumin is synthesized and secreted by *Escherichia coli*, Proc. Natl. Acad. Sci. USA 75 (12) (1978) 5936-5940

121. Kornacker, M.G., Pugsley, A.P.: The Normally Periplasmic Enzyme β-Lactamase is Specifically and Efficiently Translocated Through the *Escherichia coli* Outer Membrane when it is Fused to the Cell-Surface Enzyme Pullulanase, Mol. Microbiol. 4 (1990) 1101-1109


162. Biosep. 1 (3-4) (1990) (Special Issue: Aqueous two-phase separation systems)


168. Schustolla, D., Hustedt, H.: Affinity extraction of proteins from biomass - metal affinity extraction of D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, DECHEMA Biotech. Conf. 48 (1990) 1137-1141


177. MEDAC: pAX, a universal DNA vector system for gene expression, sequencing, and mutagenesis of gene fusion proteins, Hamburg 1991


180. DIAGEN GmbH (QIAGEN, Inc.) : The QIAexpressionist, Düsseldorf 1991

(Other sources: Invitrogen, San Diego; Novagen, Madison)


196. Soehlemann, P., Oeckl, C., Michel, H.: Expression in Escherichia coli of the genes coding for reaction center subunits from Rhodobacter sphaeroides: wild-type proteins and fusion proteins containing one or four truncated domains from Staphylococcus aureus protein A at the carboxy-terminus, Biochim. Biophys. Acta 1089 (1) (1991) 103-112


Rietveld, J., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D., Richmann,
M., Shuster, J.R.: Kluyveromyces as a host for heterologous gene expression:
expression and secretion of prochymosin, Bio/Technol. 8 (1990) 135-139

243. Heslot, H.: Genetics and Genetic Engineering of the Industrial Yeast Yarrowia

Transformation of different strains of Escherichia coli with multicopy plasmids
carrying the phoA gene causes the secretion of periplasmic alkaline phosphatase
and the accumulation of its precursor, Mol. Biol. 25 (3) (1991) 624-631

245. Edgington, S.M.: Rites of Passage: Moving Biotech Proteins Through The ER,
Bio/Technol. 10 (1992) 1413-1420

proteins carrying the HlyA transport signal during secretion by the Escherichia coli

prepropeptide portions of the Bacillus amyloliquefaciens neutral protease fused to
Bacillus subtilis α-amylase and human growth hormone during secretion in Bacillus
subtilis, J. Biotechnol. 23 (1992) 55-70


249. Wels, W., Harwerth, I.-M., Zwickl, M., Hardman, N., Groner, B., Hynes, N.E.:
Construction, bacterial expression and characterization of a bifunctional single-chain
antibody-phosphatase fusion protein targeted to the human erbB-2 receptor,
Bio/Technology 10 (1992) 1128-1132

250. WO 92/01707 (06. 02. 1992), Bionebraska, Inc. (Coolidge T.R., Wagner F., van
Heeke G., Schuster S.M., Stout J., Wylie D.: Purification directed cloning of
peptides)

Necrosis Factor Alpha in a Novel E. coli Expression System, Bio/Technol. 13 (5)
(1992) 756-761

252. Engel, H., Mottl, H., Keck, W.: A Modified Vector for the Controlled High-Level
Overproduction of Staphylococcal Protein A Fusion Proteins in the Periplasm of
Escherichia coli, Protein Expression Purification 3 (1992) 108-113

253. Rondahl, H., Nilsson, B., Holmgren, E.: Fusions to the 5' end of a gene encoding a

254. Smith, M.C.: Engineering Metal Binding Sites into Recombinant Proteins for Facile

255. WO 92/07868 (14. 05. 1992), Pharmacia AB (Enfors S.-O., Köhler K., Ljungquist C.,
Nilsson B.: A method for isolating and purifying peptides and proteins)


262. Duvoisin, R.M., Belin, D., Krisch, M.: A plasmid expression vector that permits stabilization of both mRNAs and proteins encoded by the cloned genes, Gene 45 (1986) 193-201


