1
Downstream Processing of Biotechnology Products

1.1
Introduction

Biological products are important for many applications including biotransformations, diagnostics, research and development, and in the food, pharmaceutical, and cosmetics industries. For certain applications, biological products can be used as crude extracts with little or no purification. However, biopharmaceuticals typically require exceptional purity, making downstream processing a critical component of the overall process. From the regulatory viewpoint, the production process itself defines the biopharmaceutical product rendering proper definition of effective and efficient downstream processing steps crucial early in process development. Currently, proteins are the most important biopharmaceuticals. The history of their development as industrial products goes back more than half a century. Blood plasma fractionation was the first full-scale biopharmaceutical industry with a current annual production in the 100-ton scale [1, 2]. Precipitation with organic solvents has been and continues to be the principal purification tool in plasma fractionation, although, recently, chromatographic separation processes have also been integrated into this industry. Anti-venom antibodies and other anti-toxins extracted from animal sources are additional examples of early biopharmaceuticals, also purified by a combination of precipitation, filtration and chromatography. In contrast, current biopharmaceuticals are almost exclusively produced by recombinant DNA technology. Chromatography and membrane filtration serve as the main tools for purification for these products.

Figure 1.1 shows the 2006 market share of various biopharmaceuticals. Approximately one-third are antibodies or antibody fragments [3], nearly 20% are erythropoietins, and 14% are insulins. The rest are enzymes, growth factors and cytokines [3]. Although many non-proteinaceous biomolecules such as plasmids, viruses or complex polysaccharides are currently being developed, it is likely that proteins will continue to dominate as biopharmaceuticals. Proteins are well tolerated, can be highly potent, and often posses a long half-life after administration, making them effective therapeutics. Some of these properties also make proteins potentially useful in cosmetics, although applications in this field are complicated in part by the US and European legal frameworks that do not allow the use of phar-

Protein Chromatography: Process Development and Scale-Up. Giorgio Carta and Alois Jungbauer

ISBN: 978-3-527-31819-3

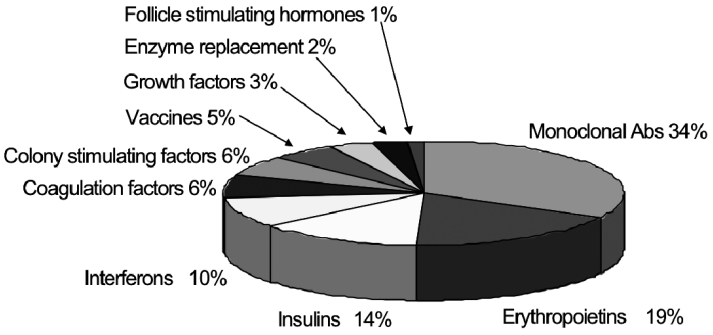


Figure 1.1 Biopharmaceuticals market share in 2006. Approximately 160 protein therapeutics have gained approval in the USA and EU. Data from La Merie Business Intelligence (www.lamerie.com).

macologically active compounds in cosmetics. Currently only a few proteins are used in this area. The most prominent one is the botulinum toxin, Botox®, used for skin care [4]. This and similar compounds are exclusively administered by physicians and thus are not considered to be cosmetics.

1.2 Bioproducts and their Contaminants

This chapter gives an overview of the chemical and biophysical properties of proteins and their main contaminants such as DNA and endotoxins. The description is not comprehensive; only properties relevant to chromatographic purification will be considered. A detailed description of the chemistry of proteins and DNA is outside the scope of this book and can be found in a number of excellent biochemistry or molecular biology texts [5, 6].

1.2.1 Biomolecules: Chemistry and Structure

1.2.1.1 Proteins

Proteins constitute a large class of amphoteric biopolymers with molecular masses ranging from 5 to 20 000 kDa, which are based on amino acids as building blocks. There are enormous variations in structure and properties within this class. Insulin, for example, a peptide with molecular mass of 5808 Da, has a relatively simple and well-defined structure. On the other hand, human van Willebrand factor, a large multimeric glycoprotein with a molecular mass of 20 000 kDa, has an extremely complex structure consisting of up to 80 subunits, each of which is 250 kDa in mass. Most proteins have a molecular mass well within these two extremes, typically between 15 and 200 kDa. Proteins are generally rather compact

molecules, yet they flexible enough to undergo substantial conformational change in different environments, at interphases, upon binding of substrates or upon adsorption on surfaces.

Proteins are highly structured molecules and their structure is generally critical to their biological function. This structure is organized into four different levels: *primary*, *secondary*, *tertiary*, and *quaternary*. The primary structure is determined by the amino acid sequence, the secondary structure by the folding of the polypeptide chain and the tertiary structure is defined by the association of multiple secondary structure domains. Finally, the quaternary structure is defined by the association of multiple folded polypeptide chains. The final result is a complex three-dimensional superstructure linked by various intra- and intermolecular interactions. Often non-amino acid elements are incorporated into a protein. Well-known examples include prosthetic groups in enzymes and iron-carrying heme groups in oxygen transport or storage proteins such as hemoglobin or myoglobin.

Primary Structure The building blocks of proteins are amino acids. During biosynthesis, following transcription and translation, these molecules are linked together via peptide bonds to form a polypeptide chain in a sequence that is uniquely determined by the genetic code. The general structure of amino acids and the formation of a peptide bond are shown in Figure 1.2. The order in which the amino acids are arranged in the polypeptide chain defines the protein's *primary structure*. Note that although amino acids are chiral molecules with L- and D-isomers, only the L-isomer is found in natural proteins. The 20 amino acids naturally found in proteins are listed in Table 1.1 In typical proteins, the average molecular mass of the amino acid components is 109 Da. Thus, the approximate molecular mass of a protein can be easily estimated from the number of amino acids in the polypeptide chain.

The peptide bond formed when amino acids are linked together has partial double bond character and is thus planar. This structure restricts rotation in the peptide chain making free rotation possible only in two out of three bonds. As a consequence, unique structures are formed depending on the particular sequence of amino acids. Certain conformations are not allowed owing to the restricted rotation, while others are energetically favored owing to the formation of hydrogen bonds and other intramolecular interactions. The amino acid side chains can be charged, polar, or hydrophobic (see Table 1.1), thereby determining the biophysical properties of a protein. The charged groups are acids and bases of differing strength or pK_a. Thus, these groups will determine the net charge of the protein

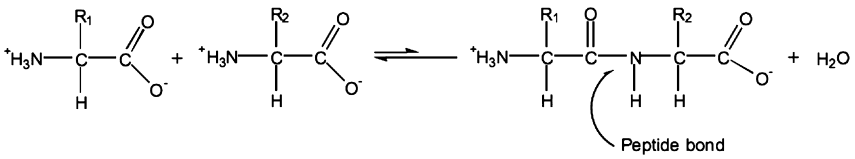


Figure 1.2 General structure of amino acids and formation of a peptide bond.

Table 1.1 The proteinogenic amino acids, including three- and one-letter codes, the structure of their R-group, relative abundance in *E. coli*, molecular mass, and pK_a of the R-group. Note that proline is a cyclic imino acid and its structure is shown in its entirety.

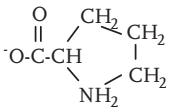
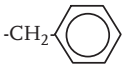
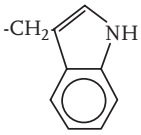
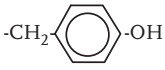
Name	3-letter code	1-letter code	R-group	Abundance in *E. coli* (%)	Molecular mass	pK_a of R-group
Hydrophobic R-groups						
Alanine	Ala	A	$-CH_3$	13.0	89	
Valine	Val	V		6.0	117	
Proline	Pro	P		4.6	115	
Leucine	Leu	L	$-CH_2-CH(CH_3)-CH_3$	7.8	131	
Isoleucine	Ile	I	$-CH(CH_3)-CH_2-CH_3$	4.4	131	
Methionine	Met	M	$-CH_2-CH_2-S-CH_3$	3.8	149	
Phenylalanine	Phe	F		3.0	165	
Tryptophan	Trp	W		1.0	204	
Polar but uncharged R-groups						
Glycine	Gly	G	$-H$	7.8	75	
Serine	Ser	S	$-CH_2OH$	6.0	105	
Threonine	Thr	T	$-CH(OH)-CH_3$	4.6	119	
Cysteine	Cys	C	$-CH_2-SH$	1.8	121	8.5
Asparagine	Asn	N	$-CH_2-C(=O)-NH_2$	11.4	132	
Glutamine	Gln	Q	$-CH_2-CH_2-C(=O)-NH_2$	10.8	146	
Tyrosine	Tyr	Y		2.2	181	10.0

Table 1.1 *Continued*

Name	3-letter code	1-letter code	R-group	Abundance in *E. coli* (%)	Molecular mass	pK_a of R-group
Acidic R-groups (negatively charged at pH~6)						
Aspartic acid	Asp	D	$-CH_2-C(=O)-O^-$	9.9	133	3.7
Glutamic acid	Glu	E	$-CH_2-CH_2-C(=O)-O^-$	12.8	147	4.2
Basic R-groups (positively charged at pH~6)						
Lysine	Lys	K	$-CH_2-CH_2-CH_2-CH_2-NH_2^+$	7.0	146	10.5
Histidine	His	H	$-CH_2-C(=CH-NH^+)-NH-CH=$	0.7	155	6.1
Arginine	Arg	R	$-CH_2-CH_2-CH_2-NH-C(=NH_2^+)-NH_2$	5.3	174	12.5

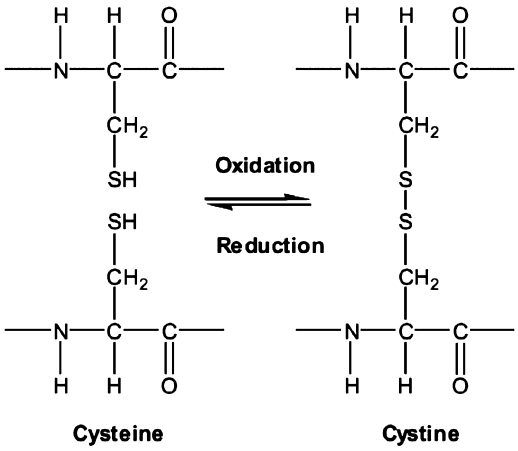


Figure 1.3 Formation of a disulfide bond upon oxidation of two cysteines.

as a function of pH. Hydrophobic side chains, on the other hand, determine the hydrophobic character of the primary structure, which plays a substantial role in determining the pattern of folding of the polypeptide chain. The amino acids cysteine and proline play particular roles. Free cysteine molecules can undergo an oxidation reaction to form disulfide bonds or bridges yielding cystine as shown in Figure 1.3. When cysteines form part of a polypeptide chain, these bridges can be either intramolecular (within the same polypeptide chain) or intermolecular to

link different polypeptide chains. On one hand, these bridges contribute to the stabilization of a protein's folded structure and on the other they can lead to the formation of covalently bonded multimeric protein structures.

The formation of disulfide bridges is generally reversible. Bonds formed in an oxidative environment can be broken under reducing conditions thus destabilizing the protein's folded structure and disrupting associated forms. This property is utilized, for example, in high-resolution analytical protein separation methods such as SDS polyacrylamide gel electrophoresis (*SDS-PAGE*) which are often carried out under reducing conditions. In this case, the resultant loss of structure and the elimination of associated forms allow the precise determination of the protein's molecular mass. Covalent chromatography utilizing the reversible formation of disulfide bonds between a protein's cysteine residues and sulfhydryl ligands bound to a surface [7] is also based on the reversible nature of these bonds and has been applied to the separation of IgG heavy and light chains.

Proline also plays a special role in defining protein structure. Proline is a cyclic imino acid and can exist in cis and trans forms. In turn, these forms influence the conformation of the polypeptide chain. In free solution, these isomeric forms are in equilibrium. However, in a polypeptide, the interconversion of these isomeric forms is often slow and can be the rate-limiting step in the establishment of folded protein structures.

Secondary Structure The polypeptide chains found in proteins do not form knots or rings and are not β-branched. However, these chains can form α-helices, β-sheets, and loops which define the protein's *secondary structure*.

α-Helices consist of a spiral arrangement of the polypeptide chain comprising 3.6 amino acid residues per turn. The helix is stabilized by intramolecular hydrogen bonds and may be hydrophobic, amphipathic or hydrophilic in character, dependent on the particular sequence of amino acids in the primary structure. Examples of such helices are given in Figure 1.4. In each case the character of the α-helix can be predicted by placing each amino acid residue in a spiral at 100 degree intervals so that there will be 3.6 residues per turn. As seen in Figure 1.4, for citrate synthase, the hydrophobic residues are dominant and uniformly distrib-

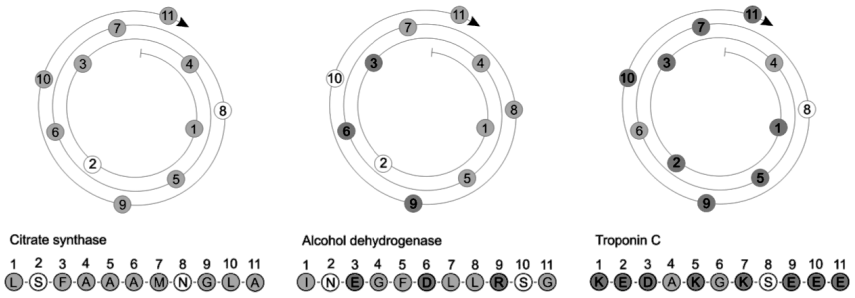


Figure 1.4 Schematic structures of hydrophobic, amphipatic, hydrophilic protein helices. Hydrophobic amino acid residues are shown in light gray, polar in white, and charged in dark gray. Based on data in [8].

uted so that the α-helix will be hydrophobic. In the last case, troponin C, the charged residues are dominant but also uniformly distributed so that the resulting helix will be hydrophilic. Finally, in alcohol dehydrogenase the hydrophobic and charged residues are non-uniformly distributed resulting in an amphipathic helix that is hydrophilic on one side and hydrophobic on the other.

β-Sheets are very stable secondary structure elements that also occur as a result of hydrogen bonding. Although one hydrogen bond makes up a free energy of bonding (ΔG) of only about $1\,\mathrm{kJ\,mol^{-1}}$, the large number of such bonds in β-sheets makes them highly stable. As seen in Figure 1.5, β-sheets have a planar structure, which can be parallel, anti-parallel, or mixed depending on the directional alignment of the polypeptide chains that form these structures. Formation of β-sheets is often observed during irreversible protein aggregation. Due to the strong intermolecular forces in these structures, vigorous denaturing agents are needed to disrupt the resulting aggregates. Urea, a strong hydrogen bond breaker, can be used for this purpose. However, the high concentrations of urea needed to disrupt the hydrogen bonding will often result in a complete destabilization and unfolding of the whole protein structure. Amyloid proteins and fibers contain a large number of β-sheets which explains in part the properties of these classes of aggregation-prone proteins.

Loops are very flexible parts of the protein and often connect other secondary structure elements with each other. For example, loops often connect the portions of a polypeptide chain that form anti-parallel areas of parallel β-sheets or form the links between different α-helical and β-sheet domains. Several types of loops have been described such as α and ω types. Loops also play a critical role in the artificial fusion of different proteins as in the case of single chain antibodies. These artificial antibodies are connected by loops that significantly contribute to the stability of the protein.

The relative number of secondary structure elements present in a protein can be measured by several spectroscopic methods including *circular dichroism* (CD) and infrared spectroscopy. CD-spectroscopy is based on the anisotropic nature of the protein. In circularly polarized light, the electric field vector has a constant length, but rotates about its propagation direction. Hence during propagation the light forms a helix in space. If this is a left-handed helix, the light is referred to as left circularly polarized, and vice versa for a right-handed helix. Due to the interaction with the molecule, the electric field vector of the light traces out an elliptical path during propagation.

At a given wavelength the difference between the absorbance of left circularly polarized (A_L) and right circularly polarized (A_R) light is

$$\Delta A = A_L - A_R \tag{1.1}$$

Although ΔA is the absorption measured, the results are usually reported in degrees of ellipticity [θ]. Molar circular dichroism (ε) and molar ellipticity, [θ], are readily interconverted by the equation

$$[\theta] = 3298.2 \cdot \Delta\varepsilon \tag{1.2}$$

A wavelength scan is used to show the content of the secondary structure of a protein and is an essential measure of integrity. It is often used either to follow

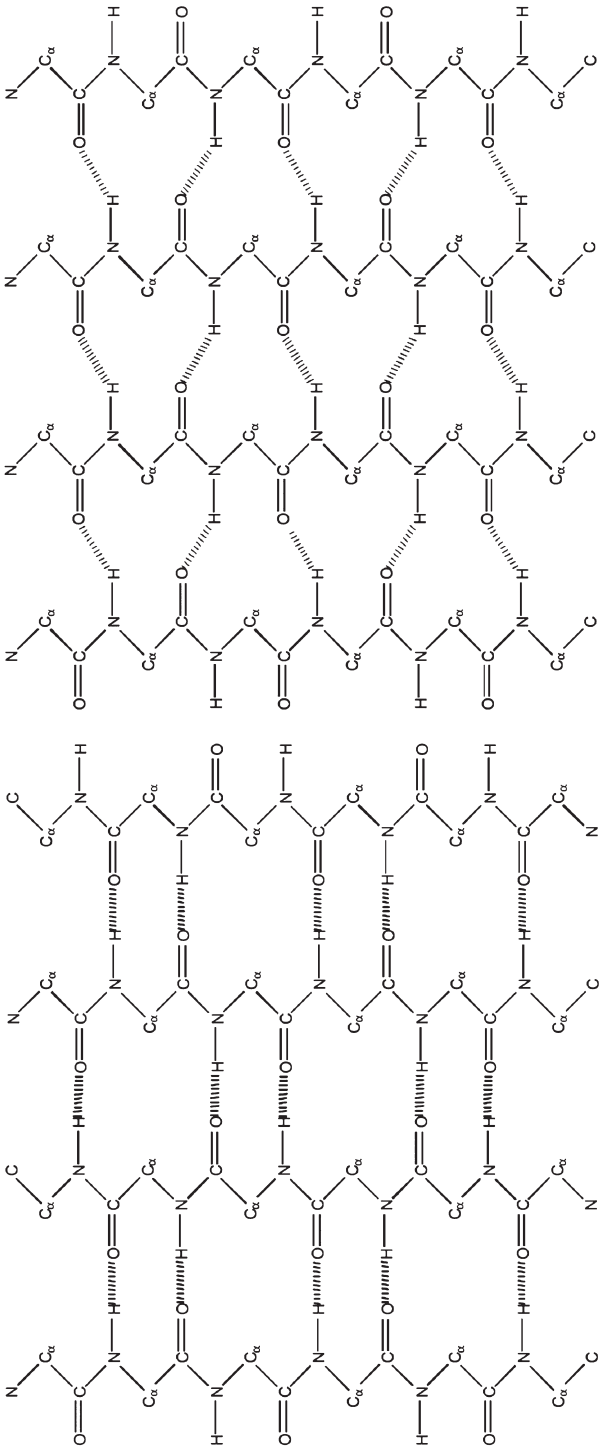


Figure 1.5 Schematic structure of parallel (left) and anti-parallel (right) β-sheets in proteins.

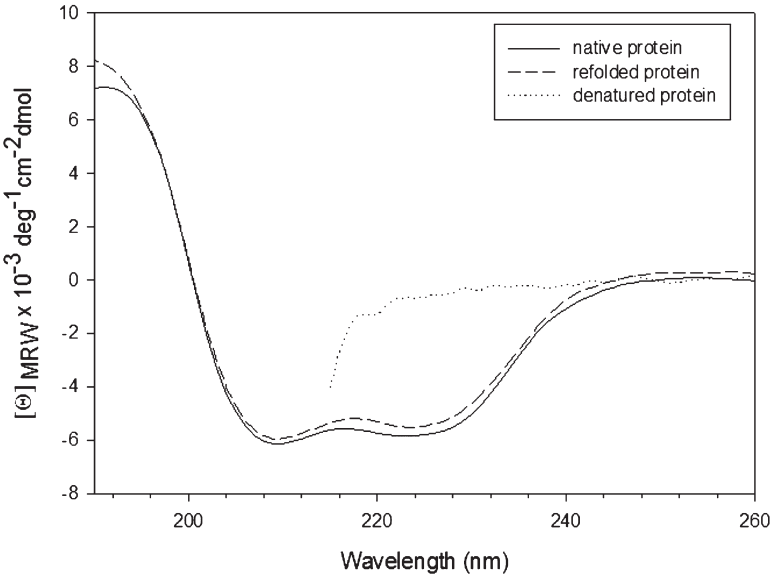


Figure 1.6 CD-Spectrum of native, refolded and unfolded α-lactalbumin.

protein refolding or to confirm the native structure of a protein (Figure 1.6). Different algorithms have been applied to determine the content of secondary structure elements based on these measurements and quantification is highly dependent on the particular algorithm used. Although CD-spectroscopy is not sufficiently sensitive to trace residual unfolded protein in a protein preparation, the method is well suited to and accepted for the study of thermally- or chemically-induced unfolding in proteins.

Attenuated total reflectance Fourier transform infrared (ATR FT-IR) spectroscopy is also used to study conformational changes in the 3D-structure of a protein *in situ*. A change in the secondary structure elements can be assessed with ATR FT-IR even in suspensions and turbid solutions. The amide I band in the spectral region from 1600 to 1700 cm^{-1} is used to evaluate structural changes (Figure 1.7). As in the case of CD, application of certain algorithms leads to the determination of the content of the secondary structure of a protein, although, again this is highly dependent on the algorithm applied. An advantage of the method is that the structure can be determined when the protein is adsorbed.

Tertiary Structure The *tertiary structure* is formed when elements of the secondary structure (α-helices, β-sheets, and loops) are folded together in a three-dimensional arrangement. Hydrophobic interactions and disulfide bridges are primarily responsible for the stabilization of the tertiary structure as exemplified by the packing of amphipatic α-helices into a four-helix bundle. In this structure, the hydrophobic residues are tightly packed in its core, shielded from the surrounding

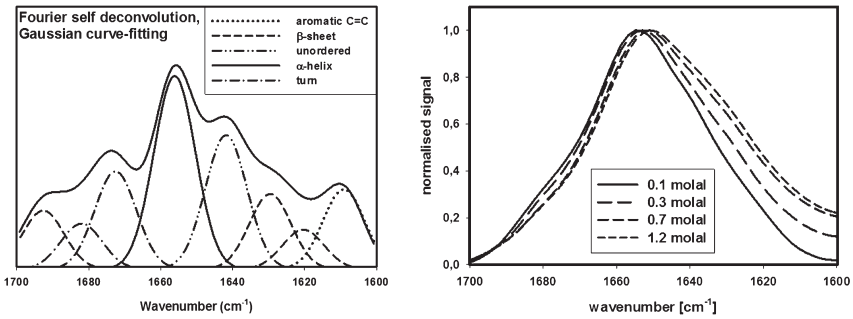


Figure 1.7 Infrared spectrum of the amide I band of a protein. The shift of the amide I band of BSA upon adsorption to the matrix during HIC with an increasing concentration of ammonium sulfate is shown on the right, indicating a significant change in secondary structure content. Reproduced from [9].

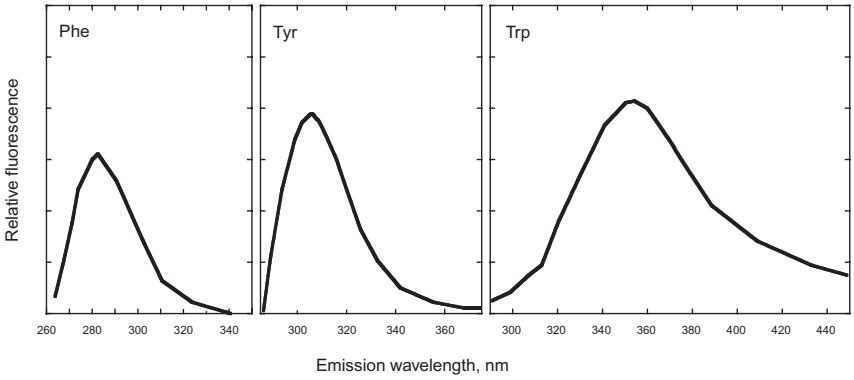


Figure 1.8 Relative fluorescence of the amino acids, Phe, Tyr and Trp. Based on data from [10].

aqueous environment, while the polar and charged residues remain exposed on its surface.

Fluorescence spectroscopy provides information about the location of the highly hydrophobic residues, tryptophan, phenylalanine and tyrosine in such folded structures. As shown in Figure 1.8, these residues have characteristic fluorescence spectra, which vary with their position in the protein structure. When these residues are exposed at the protein surface, the fluorescence maximum shifts providing an indication that unfolding has occurred. Thus, the extent of unfolding can be calculated when the fluorescence spectra of native and unfolded forms are known.

Quaternary Structure The *quaternary structure* is established when two or more polypeptide chains are associated to form a superstructure, which, in many cases, is essential for the biological function. One of the best-known examples is hemo-

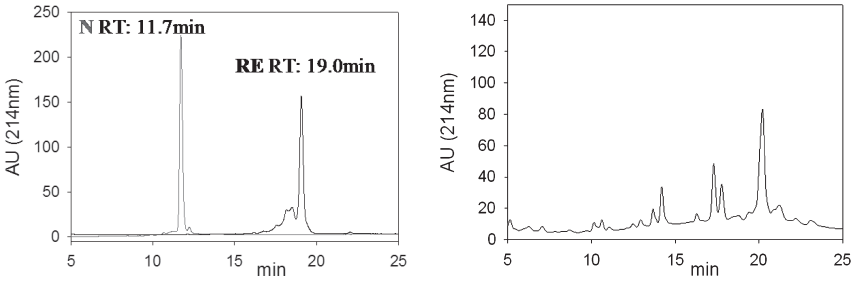


Figure 1.9 Left: retention of native and fully folded α-lactalbumin on a Vydac C4 reversed phase column containing 5 µm particles with a pore size of 30 nm. The mobile phase was a water–acetonitrile mixture. Right: separation of folding intermediates of α-lactalbumin using the same column and conditions.

globin, which consists of four polypeptide units held together by hydrogen bonding and hydrophobic interactions. In this case, the flexibility of the quaternary structure in response to oxygen binding is critical for oxygen uptake and release in the lung and capillary environments. Antibodies are another example of proteins with quaternary structures. These molecules consist of four polypeptide chains (two light and two heavy) linked together by disulfide bridges. The resulting structure is generally quite stable, allowing antibodies to circulate freely in plasma.

Folding Although individual steps in the folding pathway can be extremely rapid, the overall process of protein folding can be relatively slow. For instance the helix-coil transition and the diffusion-limited collapse of proteins occur on time scales in the order of microseconds. On the other hand, the *cis*-transproly-peptidyl isomerization is a slow reaction occurring over time scales of up to several hours. As a result, in some instances folding and the chromatography method used occur over similar time periods so that structural rearrangements can take place during separation. When folding processes are particularly slow, chromatography can be used to resolve intermediate folding variants. For example, as shown in Figure 1.9, partially unfolded proteins show different retention in reversed phase chromatography, which can be used either to analyze protein solutions during an industrial refolding process or for the preparative separation of partially unfolded forms.

Protein structures are classified into several hierarchies which include protein families and superfamilies. Dayhoff [11] introduced the term 'protein superfamily' in 1974. Currently, the term 'folds' is more commonly used to describe broad classes of protein structures. Table 1.2 shows the relative abundances of protein folds found in the PIR-International Protein Sequence Database; an excellent description of the structural hierarchies of proteins can be found on the web site: http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/description.html

Proteins have been classified into classes and folds so that common origins and evolutionary patterns can be identified. However, it should be noted that even

Table 1.2 Classes of folds found in protein databases [12].

Class of protein fold	Relative abundance
All alpha	20–30%
All beta	10–20%
Alpha and beta with mainly parallel β-sheets (α/β)	15–25%
Alpha and beta with mainly anti-parallel β-sheets with segregated α- and β-regions (α + β)	20–30%
Multi domain	<10%
Membrane and cell surface proteins	<10%
Small proteins (dominated by cofactors or disulfide bridges)	5–15%

proteins belonging to the same class may behave differently since even the substitution of a single amino acid can result in large variations in biophysical properties.

Post-Translational Modifications Post-translational modifications are often critical to a protein's biological function and can dramatically impact downstream processing. Post-translational modifications occur after the primary structure is formed and are highly cell specific. They can also vary with the physiological status of the cell. The latter can vary, for example, when the cells are deprived of a certain nutrient or find themselves in a low oxygen environment. Further modifications can also occur following expression. As a result, many protein-based biopharmaceuticals are highly heterogeneous and their biological and pharmacological activity is often greatly influenced by the production process. Difficulties encountered in fully characterizing the corresponding broad range of molecular diversity often require that protein pharmaceuticals be defined by the process by which they are produced rather than as uniquely defined molecular entities. While considerable effort is being devoted to developing so-called 'well-characterized biologicals', for which molecular qualities, rather than processing define the product, current regulations continue to define biologicals strictly by their manufacturing process.

Post-translational modifications often represent a productivity bottleneck. At high expression rates, post-translational modifications are often altered or become incomplete when the cell's ability to perform these transformations lags behind the protein translation machinery. The result is the expression of additional protein variants, with potentially varying biological activity, stability, and biophysical properties such as solubility, charge, hydrophobicity, and size. Thus, improved fermentation or cell culture titers often have to be balanced against the increased heterogeneity of the product formed.

More than 200 types of post-translational modifications have been described in the literature. Table 1.3 summarizes the most relevant ones. The individual molecular entities produced by such modifications are called isoforms.

Two especially relevant post-translational modifications are glycosylation and deamidation since both produce changes that can influence the chromatographic properties of the protein. Thus, chromatography can be used as a tool to separate

Table 1.3 Examples of post-translational modifications of proteins. Data from http://www.expasy.ch/tools/findmod/findmod_masses.html.

Modification	Characteristics	Average mass difference
Glycosylation	O-linked oligosaccharides bound to Ser and Thr; N-linked oligosaccharide bound to Asp	Varies with number of sugar moieties; up to several thousands
Phosphorylation	Ser, Thr and Tyr, a phosphoester is formed, typical modification of allosteric proteins involved in regulation (signal transduction)	79.9799
Sulfation	Addition of sulfate to Arg and Tyr, a C-O-SO_3 bond	80.0642
Amidation	Addition of -NH_2 to C-terminus	–0.9847
Acetylation	Addition of CH_3CO- to N- or C-terminus	42
Hydroxylation	Addition of -OH to Lys, Pro and Phe	16
Cyclization	Formation of Pyroglutamate at N-terminal Glu	–0.9847
Complexation of metals	Cys-CH_2-S-Fe complexes in Ferrodoxins. Selenium-complexes with Cys and Met Copper-complexes with backbone of peptide bond	
Halogenation	Iodination and bromination of Tyr (3-chloro, 3-bromo)	34, 78
Desmosin formation	Desmosin is formed by condensation of Lys, frequent in elastin	–58
γ-Carboxylation	In prothrombin and blood coagulation factor VII	44.0098
Hydroxyproline	Hydroxyproline formation in collagen responsible for mechanical stability	15.9994
Adenylation	Tyr residue of glutamine synthetase is adenylated	209
Methylation	Addition of methyl group to Asp, Gln, His, Lys und Arg of flagella protein	14.0269
Deamidation	Asn und Gln are susceptible; both biological and processing deamidation are observed	0.9847

the corresponding isoforms. Glycosylation is the addition of carbohydrate molecules, either simple sugars or complex oligosaccharides, to the protein molecule. Glycosylation renders the protein more hydrophilic and thus more soluble. Additionally, however, since the terminal carbohydrates of such oligosaccharides are often neuraminic acids (generally known as sialic acid) which are negatively charged above pH 3, glycosylation also influences the net charge and isoelectric point of the protein. As a result, chromatographic separations based on the different charge of the glycovariants are possible.

As an example, Figure 1.10 shows the *isoelectric focusing* (IEF) gel separation of recombinant human erythropoietin (rhEPO – currently the second largest seller in the biopharmaceutical industry). As can be seen in this figure, the starting material

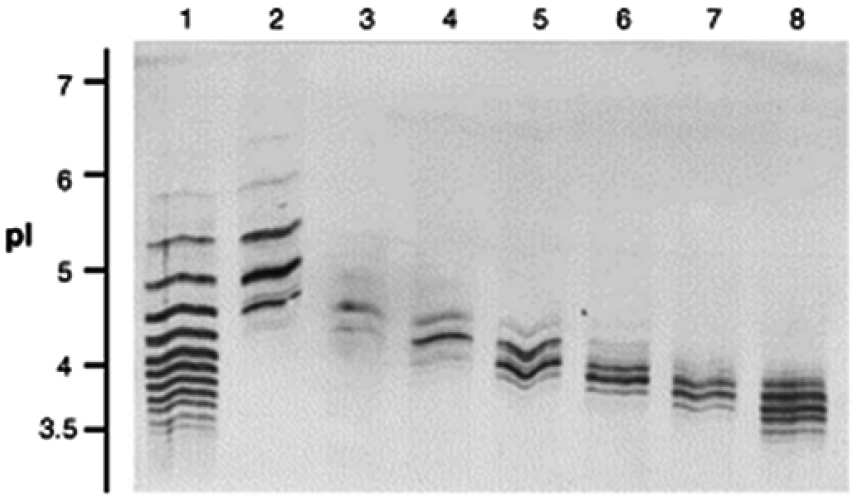


Figure 1.10 Isoelectric focusing (IEF) of rhEPO. Fractions obtained by DEAE–Sephacel chromatography: (1) starting material; (2) unadsorbed material; (3) material eluted with 0.015 M; (4) 0.03 M; (5) 0.06 M; (6) 0.15 M; (7) 0.35 M and (8) 1 M NaCl. Reproduced from [13] with permission.

contains multiple variants with isoelectric points between 3.5 and 5.5. Loading the starting material on an anion exchange column and eluting with increasing salt concentrations results in eluted fractions that have substantially reduced heterogeneity. Later eluting fractions contain more acidic variants with lower isoelectric points. These variants are more negatively charged and elute only at higher salt concentrations from the positively charged anion exchanger. rhEPO is highly glycosylated and the glycovariants have different bioactivity. Thus, control of the glycosylation pattern and, in some cases, separation of certain undesirable variants is needed to maintain a consistent product quality.

Deamidation can also have dramatic effects both on bioactivity and chromatographic behavior. Deamidation involves the chemical transformation of asparagine and glutamine, which are uncharged polar amino acids, into aspartic acid and glutamic acid respectively, both of which are negatively charged at pH values above 4. Deamidation of asparagine residues is observed more frequently than that of glutamines, but the process is highly dependent on the location of these residues in the protein structure. Surface exposed residues tend to be most affected, while those buried within the protein core are usually partially protected. Deamidation is generally facilitated by higher pH values and higher temperatures and occurs via the mechanism illustrated in Figure 1.11. In this process, an amino group is cleaved off from asparagine forming an L-cyclic imide intermediate. This intermediate is generally unstable and is further converted into L-aspartyl and L-iso-aspartyl peptides. Both introduce negative charge and lower the isoelectric point of the protein. It should be noted that the unstable L-cyclic amide can also undergo

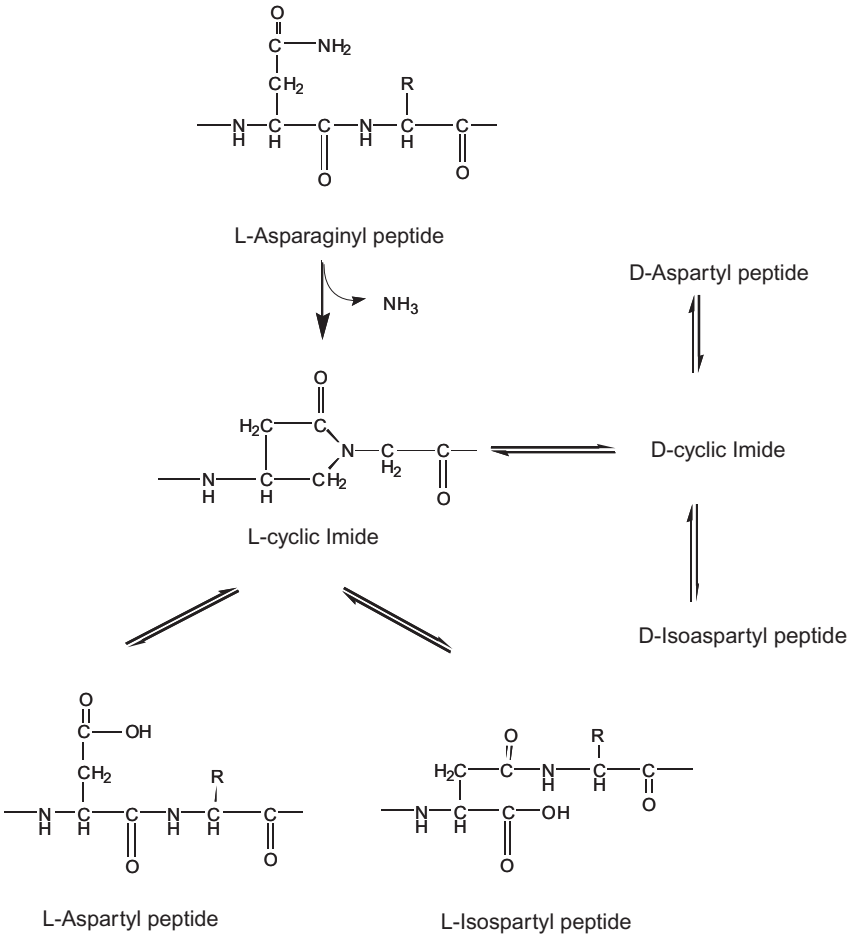


Figure 1.11 Mechanism of deamidation of proteins for asparagine residues.

racemization forming a D-cyclic amide, which is further converted into D-aspartyl and D-isoaspartyl peptide. The net result is the introduction of D-amino acids into a protein. Removal of deamidated variants is often an important task since these variants can have different bioactivity and their removal is a challenge for downstream processing. Separation by ion-exchange chromatography is possible but often difficult since the net charge difference between native and deamidated forms can be small, resulting is low selectivity.

1.2.1.2 **Oligonucleotides and Polynucleotides**

Oligonucleotides and polynucleotides are either contaminants or may constitute the product. For example, in the production of plasmid DNA for gene therapy applications, genomic DNA is a contaminant [14]. Conversely, in the production of protein pharmaceuticals, both genomic and plasmid DNA are contaminants.

Polynucleotides are present in the cell either as *deoxyribonucleic acid* (DNA) or as *ribonucleic acid* (RNA). DNA or RNA encode genetic information. In humans, animals or plants DNA is the genetic material, while RNA is transcribed from it. In some other organisms such as RNA viruses, RNA is the genetic material and, in reverse fashion, the DNA is transcribed from it. The building blocks of these molecules are nucleotides, which, in turn, are composed of a phosphate group, a sugar group, and a nitrogenous nucleoside group. Nucleotides are thus rather hydrophilic and negatively charged because of the acidic phosphate group. In DNA, the nucleotides are arranged in a double-stranded helical structure held together by weak hydrogen bonds between pairs of nucleotides. The molecule resembles a twisted 'ladder', where the sides are formed by the sugar and phosphate moieties, while the 'rungs' are formed by the nucleoside bases joined in pairs with hydrogen bonds.

There are four nucleotides in DNA, each containing a different nucleoside base: adenine (A), guanine (G), cytosine (C), or thymine (T). Base pairs form naturally only between A and T and between C and G so that the base sequence of each single strand of DNA can be simply deduced from that of its partner strand.

RNA is similar to DNA in structure but contains ribose instead of deoxyribose. There are several classes of RNA molecules including messenger RNA, transfer RNA and ribosomal RNA. They play a crucial role in protein synthesis and other cell activities. miRNAs are global regulators of gene expression. miRNAs are non-coding double-stranded RNA molecules comprising 19 to 22 nucleotides that regulate gene expression at the post-transcriptional level by forming a conserved single-stranded structure and showing antisense complementarity that was identified initially in the nematode *Caenorhabditis elegans*.

DNA and RNA are chemically very stable molecules unless DNAse or RNAse enzymes are present. In presence of these ubiquitous enzymes the polynucleotides are rapidly degraded. Polynucleotides are also very sensitive to mechanical shear. Upon cell lysis, DNA and RNA are released into the culture supernatant and dramatically alter the viscosity of fermentation broths as a result of their size and filamentous structure.

Genomic DNA present in the nucleus of eukaryotic organisms is always associated with very basic proteins known as histones. Plasmid DNA, on the other hand, is present in the cytoplasm of prokaryotic organisms and is histone-free but exists in different physical forms including supercoiled, circular, linear, and aggregated as illustrated in Figure 1.12.

These forms differ in size providing a basis for separation by gel electrophoresis or by size exclusion chromatography. Polynucleotides are negatively charged over a wide range of pH due to the exposed phosphate groups. Thus, they are strongly bound by positively charged surfaces. As a result, their removal in downstream processing is conveniently and efficiently carried out with anion-exchange resins or with positively charged membranes.

1.2.1.3 **Endotoxins**

Endotoxins, also known as *pyrogens*, are components of the cell wall of Gram-negative bacteria. They are continuously excreted by bacteria and are ubiquitous

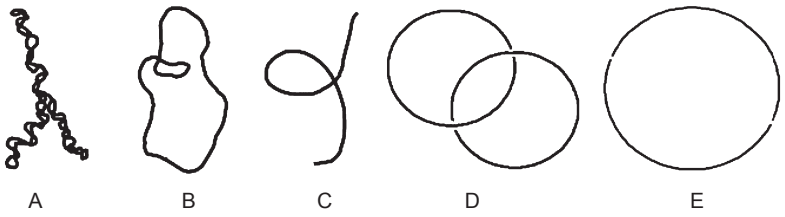


Figure 1.12 Different physical forms of plasmid DNA. (A) The linear strand is twisted to a supercoil; the supercoiled form has the highest transformation efficiency and is the predominant form in therapeutic plasmids. (B) When one strand is nicked then an open circular form is generated (C) and with the cleavage of the double strand the linear form is generated. (D) Two circular forms generate a catanane or a (E) concatemer.

in bioprocessing. Endotoxins are extremely toxic when they enter the bloodstream and humans are among the most endotoxin-sensitive organisms. Thus their almost complete removal from the finished product is required. As shown in Figure 1.13, endotoxins are lipopolysaccharides comprising a lipid A moiety, a core region, and an O- or S-antigen. The lipid A moiety is the most conserved component and is found in all endotoxins. This is also the part of the molecule responsible for toxicity. The O- or S-antigen is highly variable and strain specific. The size and structure also depends on the growth conditions.

Endotoxins target the immune responsive cells such as macrophages, monocytes, endothelial cells, neutrophils and granulocytes. They induce the expression of interleukins, tumor necrosis factor, colony-stimulating factor, leukotrienes and oxygen radicals in these cells. As a consequence of the presence of endotoxins in the bloodstream, the patient develops tissue inflammation and fever, drop in blood pressure, shock, palpitations, a decrease in vessel permeability, respiratory complications, and even death. The same symptoms occur with severe bacterial infection, so-called septic shock. Severe hepatic toxicity and hematological disorders have been observed to occur in humans in response to as little as 8 ng of endotoxins per kg body weight. In contrast, endotoxins are much less toxic to many animals. For example, the LD_{50} is as high as 200–400 µg/animal in mice. For parenteral biopharmaceuticals the threshold level for intravenous applications is 5 *endotoxin unit*s (EU) per kg body weight per hour. EU defines the biological activity of endotoxins with 1 EU corresponding to 100 pg of the EC-5 standard endotoxin or 120 pg of the endotoxin derived from the *E. coli* strain O111:B4. The detection of endotoxin is difficult and is carried out using bioassays. In the past rabbits have been used for this purpose. This time-consuming test has been replaced by the so-called *limulus amoebocyte lysate* (LAL) test, which uses the hemolymph of the horseshoe crab. LAL coagulates in the presence of minute amounts of endotoxins (see Figure 1.14) forming the basis for assays with endotoxin detection limits as low as of 10 pg ml^{-1}. General guidelines are described in the *United States Pharmacopeia* (USP) in Chapter 79 on pharmaceutical *compounding and sterile preparation*s (CSP).

Table 1.4 provides a summary of the typical endotoxin content of various solutions. Endotoxins are present in large concentration in protein solutions derived

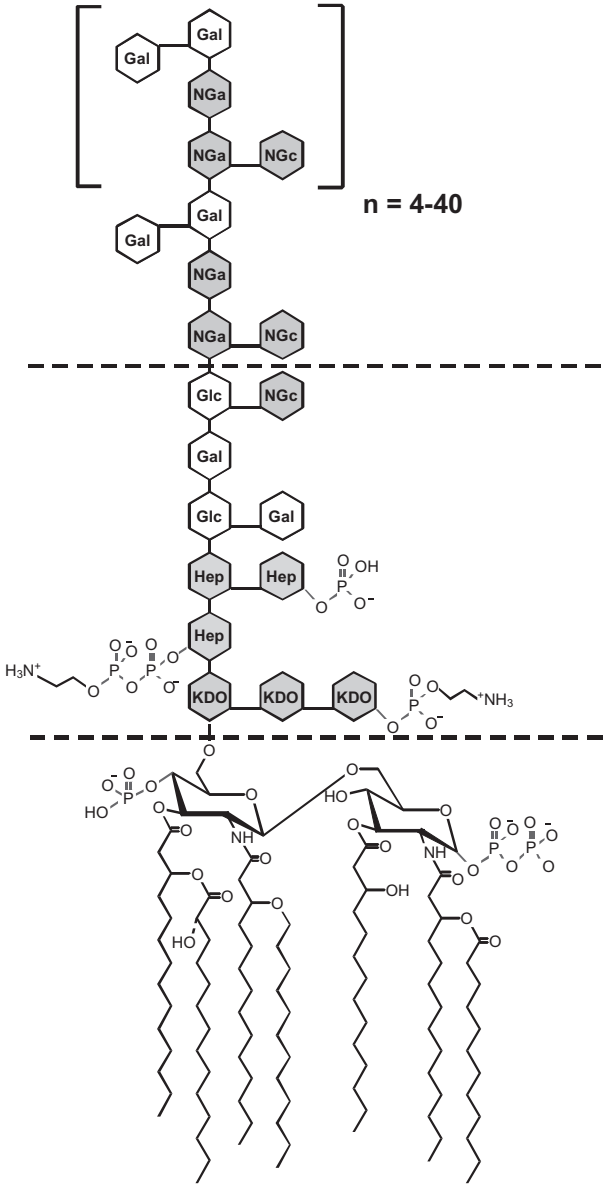


Figure 1.13 Chemical structure of endotoxins. Reproduced from [15] with permission.

from bacterial fermentations, but can also be present as adventitious agents in many other systems.

In the industrial production of pharmaceuticals for parenteral use, special care is used to prevent endotoxin contamination. For example, endotoxin-free water used in the preparation of culture media and chromatography buffers, is re-

Figure 1.14 Coagulation test using Amoebocyte lysate for detection of endotoxins. The lysate forms a gel in the presence of endotoxins. Reproduced with permission of Associates of Cape Cod, Inc.

Table 1.4 Typical endotoxin concentrations in various solutions of crude and purified proteins [15].

Protein source	Solution	Endotoxin (EU ml^{-1})
Proteins from high-cell-density culture of *E. coli* TG:pλFGFB	Supernatant after homogenization	>>1 000 000
Proteins from shaking-flask culture of *E. coli*	Culture filtrate	70 000–500 000
Murine IgG1 from cell culture	Culture filtrate	97
Whey processed from milk of local supermarket	Supernatant after acid milk precipitation	9900
Commercial preparation of BSA	Reconstituted lyophisate at a concentration of 1 mg ml^{-1}	50 (Supplier I) 0.5 (Supplier II)

circulated at high temperature in order to avoid bacterial growth and the consequent formation of endotoxins. Although endotoxins are heat stable, they are destroyed at alkaline pH. Thus, cleaning processing equipment, tanks, membranes, and chromatography media with a sodium hydroxide solution is generally required to assure complete removal of these contaminants.

1.2.2 Biomolecules: Physiochemical Properties

1.2.2.1 UV Absorbance

The concentration of a protein in solution is often quantified by UV absorbance which is primarily due to absorption by the aromatic amino acids tyrosine, tryptophan, and phenylalanine and the disulfide bridges. The wavelength absorbance maxima and corresponding extinction coefficients for these components are summarized in Table 1.5.

Because of the strong absorbance of tryptophan, absorption maxima for proteins are typically around 280 nm and this wavelength is most frequently used for quantitative determinations. According to the Lambert–Beer law, the absorbance of a protein solution at a given wavelength defined as

Table 1.5 Absorbance characteristics of aromatic amino acids and disulfide bridges.

Amino acid	λ_{max} (nm)	ε_m^{max} ($M^{-1}cm^{-1}$)	ε_m^{280} ($M^{-1}cm^{-1}$)
Tryptophan	280	5500	5600
Tyrosine	275	1490	1400
Phenylalanine	258	200	
Disulfide bridge			134

Table 1.6 Representative values of the specific absorbance of proteins at 280 nm in a cuvette of 1-cm length at a concentration of 1 mg ml^{-1} and the molar extinction coefficient. Molar extinction coefficients from [16].

Protein	Molecular mass	Number of amino acids Trp-Tyr-Cys	Mass extinction coefficient E_{1cm}^{280} (ml $mg^{-1}cm^{-1}$)	Molar extinction coefficient ε_m^{280} ($M^{-1}cm^{-1}$)
Immunoglobulin G[a)]	155 000	Varies with subclass and individual antibody	≈1.4	
Chymotrypsinogen	50 600	8-4-5	2.0	50 600
Lysozyme (hen egg white)	14 314	6-3-4	2.73	37 900
β-Lactoglobulin	18 285	2-4-2	0.95	17 400
Ovalbumin (chicken)	42 750	3-10-1	0.74	32 000
Bovine serum albumin	66 269	2-20-17	0.68	45 000
Human serum albumin	66 470	1-18-17	0.58	39 800

a) May vary with recombinant IgG, when variable domains contain an excess of aromatic amino acids.

$$A = -\log\frac{I}{I_0} \tag{1.3}$$

is linearly related to the molar concentration of the analyte, c, by the following equation:

$$A = \varepsilon_m l c \tag{1.4}$$

where I_0 is the incident light, I is the light transmitted through the solution, l is the length of the light path through the solution and, ε_m is the specific molar absorbance or extinction coefficient. The validity of Equation 1.4 is generally limited to relatively dilute solutions and short light paths, for which A is less than 2. At higher values, the ratio of transmitted and incident light becomes too small to permit a precise determination. Thus, quantitative determinations of concentrated protein solutions require dilution or very short light paths.

As shown in Table 1.6, the specific absorbance of typical proteins varies with the relative content of the aromatic amino acids Trp and Try and, to a lesser extent, of the disulfide bridges. Since the relative content varies for different proteins, an empirical determination is needed for exact quantitative determinations.

Alternatively, the molar absorption coefficient can be estimated with relative accuracy as the linear combination of the individual contributions of the Trp and Tyr residues and of the disulfide bridges according to the following equation:

$$\varepsilon_m^{280}(\mathrm{M^{-1}cm^{-1}}) = 5500 \times n_{Trp} + 1490 \times n_{Tyr} + 125 \times n_{SS} \quad (1.5)$$

where n_{Trp}, n_{Tyr}, and n_{SS} are the numbers of its Trp, Tyr residues and disulfide bonds, respectively.

It should be noted that nucleic acids have an absorbance maximum at 260 nm and can interfere substantially with protein determinations at 280 nm. Thus, when nucleic acids are simultaneous present in solution, corrections must be made in order to determine protein concentration from absorbance values at 280 nm.

The peptide groups of proteins absorb light in the 'far-UV' range (180–230 nm) and very high absorbance values are observed in this region even for very dilute conditions. As a result, detection in analytical chromatography is often carried out at 218 nm, where absorbance is about 100 times greater.

Proteins with additional chromophores either absorb in the near-UV or visible wavelength range. Typical examples are the iron-containing proteins such as hemoglobin, myoglobin and transferrin which are red in color, or Cu-Zn superoxide dismutase which is green.

Nucleic acids show strong absorbance in the 240–275 nm region due to the π-π* transitions of the pyrimidine and purine nucleoside rings. Polymeric DNA and RNA absorb over a broad range with a maximum near 260 nm. The specific mass extinction coefficient of DNA E_{1cm}^{260} is 20 ($\mathrm{ml\,mg^{-1}\,cm^{-1}}$). The purity of DNA is estimated by the ratio of absorbance at 260 and 280 nm. For pure double-stranded DNA and RNA the ratio E^{260}/E^{280} is between 1.8 and 2.0. The measurements are more reliable at alkaline pH. In contrast to proteins, the absorbance of nucleic acids is fairly sensitive to pH, and decreases at lower pH values [17].

1.2.2.2 Size

Solutions and suspensions found in downstream processing of biotechnology products contain molecules and particles with a broad range of sizes as illustrated in Table 1.7. Globular proteins are in the range of 3–10 nm, while nucleic acids can be much larger. Therapeutic plasmids are in the range of 100 nm. Virus and virus-like particles are in the range of 50 nm to 400 nm, while cells are in the micrometer range.

While cells and cell debris are easily separated by centrifugation due to their high sedimentation velocity (Table 1.7), proteins and nucleic acids require more sophisticated methods such as chromatography and membrane filtration. Separation of proteins by ultracentrifugation is only carried out for analytical purposes since extremely high rotation rates (as high as 50 000 rpm) are needed.

The sizes given in Table 1.7 are for folded globular proteins. In this state, native protein structures are quite dense (mass density ~1.4 $\mathrm{g\,cm^{-3}}$) and are spherical or ellipsoid in shape. However, denatured, fibrous, rod, or disk shaped proteins deviate from these compact shapes. In these cases, the size of the proteins and other macromolecules is often described by other parameters which include the radius of gyration, r_g, the hydrodynamic radius, r_h, the radius established by rotat-

Table 1.7 Categories of bioproducts and their sizes.

Category	Example	M_r (Da)	Size	Sedimentation velocity (cm h^{-1})
Small molecules	Amino acids	60–200	0.5 nm	
	Sugars	200–600	0.5 nm	
	Antibiotics	300–1000	1 – nm	
Macro molecules	Proteins	10^3–10^6	3–10 nm	$<10^{-6}$
	Nucleic acids	10^3–10^{10}	2–1000 nm	
Particles	Viruses		50–500 nm	$<10^{-3}$
	Bacteria		1 µm	0.02
	Yeast cells		4 µm	0.4
	Animal cells		10 µm	2

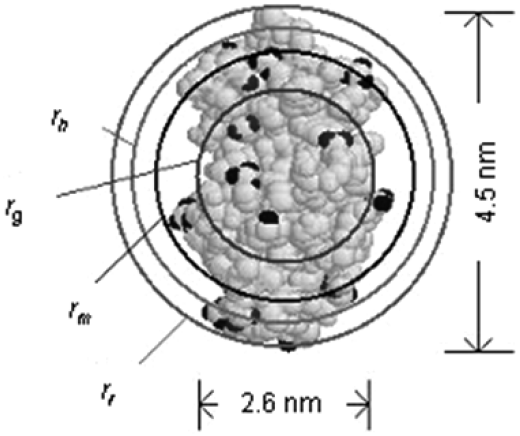


Figure 1.15 Lysozyme with a size of 2.6 × 4.5 nm is an ellipsoid shaped molecule. The molecular mass is 14.7 kDa, the mass density is 1.37 g/cm^3. r_m is the equivalent radius of a sphere with the same mass and particle specific volume as lysozyme. r_r is the radius established by rotating the protein about its geometric center.

ing the protein about its geometric center, r_r, and the radius, equivalent to a sphere with the same mass and density as the actual molecule, r_m.

Figure 1.15 illustrates these four different parameters for lysozyme. Note that the first two, r_h and r_g, can be obtained from direct biophysical measurements, while the last two, r_m and r_r, can only be inferred from a knowledge of the actual protein structure.

The radius of gyration can be measured by static light scattering. This is often carried out in conjunction with *size exclusion chromatography* (SEC) thus enabling protein mixtures to be analyzed. A general relationship exists between the radius of gyration and the amount of light scattered, which is directly proportional to the product of the weight-average molar mass and the protein concentration. Accordingly,

$$\frac{k \cdot c}{R(\theta)} = \frac{1}{M_w P(\theta)} + 2A_2 c \tag{1.6}$$

where $R(\theta)$ is the excess intensity of scattered light at a certain angle (θ), c is the sample concentration, M_w the weight-average molar mass, A_2 the second viral coefficient, k is an optical parameter equal to $4\pi n^2 (dn/dc)^2 / (\lambda_0^4 N_A)$. n is the solvent refractive index and dn/dc is the refractive index increment, N_A is Avogadro's number, λ_0 is the wavelength of scattered light in vacuum. The function $P(\theta)$ describes the angular dependence of scattered light. The expansion of $1/P(\theta)$ to first order gives:

$$\frac{1}{P(\theta)} = 1 + \left(\frac{16\pi^2}{3\lambda^2}\right) r_g^2 \cdot \sin^2\left(\frac{\theta}{2}\right) + f_4 \sin^4\left(\frac{\theta}{2}\right) + \ldots \tag{1.7}$$

At low angles the angular dependence of light scattering depends only on the mean square radius r_g^2 (alternatively known as the radius of gyration) and is independent of molecular conformation or branching.

The hydrodynamic radius can be related to the protein translational diffusion coefficient, D_0, using the Stokes–Einstein equation:

$$r_h = \frac{k_b T}{6\pi\eta D_0} \tag{1.8}$$

where k_b is the Boltzmann constant, T is the absolute temperature, and η is the solution viscosity. Accordingly, r_h represents the radius of a sphere with the same diffusion coefficient as the actual protein. D_0 can be conveniently determined by *dynamic light scattering* (DLS) also in conjunction with SEC in the case of mixtures. DLS is based on the fluctuations or Brownian motion of a molecule, which in turn cause fluctuations in the intensity of scattered light. The corresponding signal change with time can be described by an autocorrelation function. For small angles or Q-values the correlation function $C(t)$ can be expressed by a single exponential term that allows the determination of D_0 from the following equation:

$$C(t) = A_1 + A_2 e^{-2DQ^2 t} \tag{1.9}$$

where A_1 and A_2 are the baseline at infinite delay and the amplitude at zero delay of the correlation function, respectively. Tanford [18] has shown that the hydrodynamic radius of a globular protein can be related to its molecular mass, M_r, by a simple relationship. For practical calculations, the following equation provides reasonable values:

$$r_h \approx 0.081 \times (M_r)^{\frac{1}{3}} \tag{1.10}$$

where r_h is in nm.

An alternative, commonly-used method for the determination of protein size is *size exclusion chromatography* (SEC). Molecules of different sizes do not all penetrate the pores of a SEC medium to the same degree thus leading to varying retention in the column. The SEC column can be calibrated with protein standards of known molecular mass allowing the size of an unknown protein to be estimated from its retention.

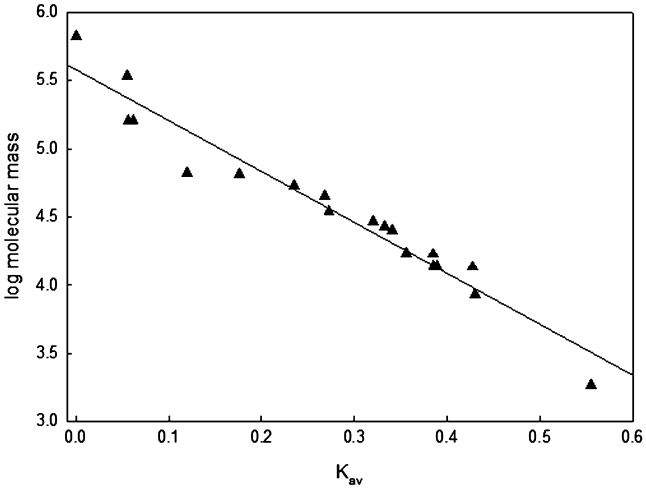


Figure 1.16 Calibration of a size exclusion column (Superdex 75, GE Healthcare, Uppsala, Sweden) with a set of reference proteins (molecular mass in parenthesis): (1) thyroglobulin (669000), fibrinogen (340000), glucose oxidase (160000), IgG (160000), bovine serum albumin (66430), hemoglobin, (64500), trisosephophate isomerase (53200), ovalbumin (45000), lectin (35000), carbonic anhydrase (29000), subtilisin (27000), chymotrypsinogen (25000), myoglobin (17000), calmodulin (16800), ribonuclease A (13700), ribonuclease S (13700), cytochrome c (13600), ubiquitin (8600), and Pep6His (1839). The symbol K_{av} is used in lieu of K_D.

An example is shown in Figure 1.16. The distribution coefficient, K_D is defined as follows:

$$K_D = \frac{V_R - V_0}{V_t - V_0} \tag{1.11}$$

where V_R is the retention volume, V_t the total column volume, and V_0 the extra-particle void volume. The latter is determined empirically from the retention of a compound sufficiently large to be completely excluded from the pores of the chromatography matrix. Blue Dextran, a 2000-kDa molecular mass dextran labeled with a blue dye, is often used for this purpose. When K_D is plotted versus the logarithm of molecular mass an almost linear relationship is obtained for standard proteins.

Other methods for the determination of the molecular size of proteins are SDS-polyacrylamide gel electrophoresis (SDS-PAGE) which provides information about the molecular mass, ultracentrifugation which provides information regarding the hydrodynamic radius, and other scattering techniques such as *small angle X-ray scattering* (SAXS).

1.2.2.3 **Charge**

Proteins are amphoteric molecules with both negative and positive charges, which stem from the side chains of acidic and basic amino acids (Table 1.1) and from the amino and carboxyl terminus of each polypeptide chain. The latter have pK_a values

around 8.0 and 3.1, respectively. Modification of amino acid side chains may substantially contribute to the charge of a protein. Important examples are glycosylation with sialic acid which occurs, for example, at N-glycosylation sites in antibodies or erythropoietin, and deamidation of asparagine and glutamine residues. Both post-translational modifications make proteins more acidic and thus more highly negatively charged. In many cases they also affect the *in-vivo* half-life of the protein, so that their control can be an important goal in downstream processing.

The net charge of a protein depends on the number of ionizable amino acid residues and their pK_a values. The protonation of these residues changes with pH according to the following equations for acidic and basic residues, respectively:

$$K_a = \frac{[R^-][H^+]}{[RH]} \tag{1.12}$$

$$K_a = \frac{[R][H^+]}{[RH^+]} \tag{1.13}$$

where the brackets indicate thermodynamic activities. In logarithmic form, we obtain:

$$\log\frac{[R^-]}{[RH]} = pH - pK_a \tag{1.14}$$

$$\log\frac{[R]}{[RH^+]} = pH - pK_a \tag{1.15}$$

where p indicates $-\log_{10}$. From these equations it is obvious that acidic residues are completely deprotonated and thus negatively charged at pH values that are two units higher than their pK_a. Conversely, basic amino acids are completely protonated and thus positively charged at pH values that are two units below their pK_a. Based on the pK_a values shown in Table 1.1, we can see that in practice, at the neutral pH values typically encountered in bioprocessing, all acidic residues in protein are negatively charged while all basic residues are positively charged. Histidine however, is an exception to this rule. Its pK_a is near neutral; thus, under typical processing conditions, this residue will be charged to an extent that depends on the exact value of pH.

At a particular pH, known as the isoelectric point or pI, the protein net charge becomes zero with an exact balance of positively and negatively charged residues. Knowing the pK_a values of the side chains and the primary sequence the net charge and the theoretical isoelectric point can be calculated from Equations 1.12 and 1.13. An example is shown in Figure 1.17 for lysozyme. The calculation is only approximate because activity coefficients were neglected and the pK_a values were assumed to be equal to those of the free amino acids. This is likely incorrect since the microenvironment where the individual residues are actually found in the protein structure has a significant effect. Nevertheless, the agreement between the theoretical net charge and that determined experimentally as a function of pH is remarkable. The more significant deviations in this case occur for native lysozyme at low pH, but largely disappear when this protein is denatured, suggesting that the discrepancy arises because some of the acidic residues may be

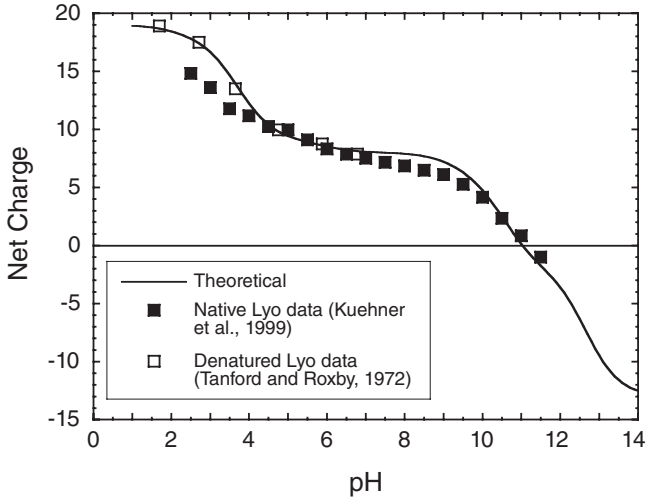


Figure 1.17 Net charge of lysozyme in denatured and native forms compared to theoretical calculation. Data from [19, 20].

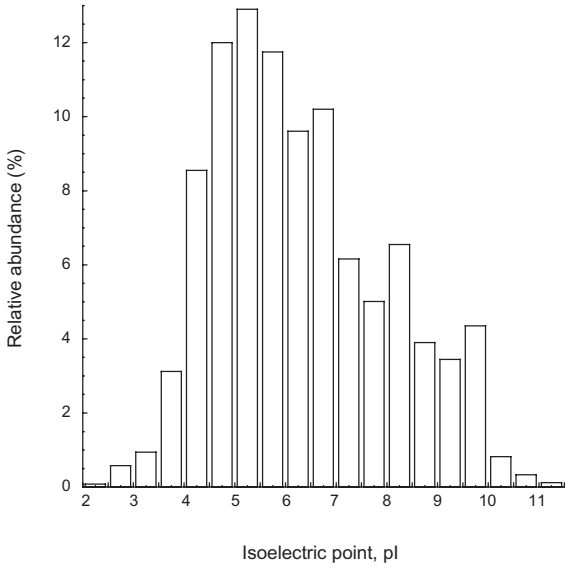


Figure 1.18 Distribution of isoelectric points of proteins. Data from [21].

partially buried in the folded structure. As shown in Figure 1.17, the pI of lysozyme is around 11 and this is in agreement with IEF measurements. Around neutral pHs, this protein has a high net positive charge with a plateau region where the charge is only slightly affected by pH. Such conditions would be conducive to a robust adsorption process for the capture of lysozyme using a cation exchanger.

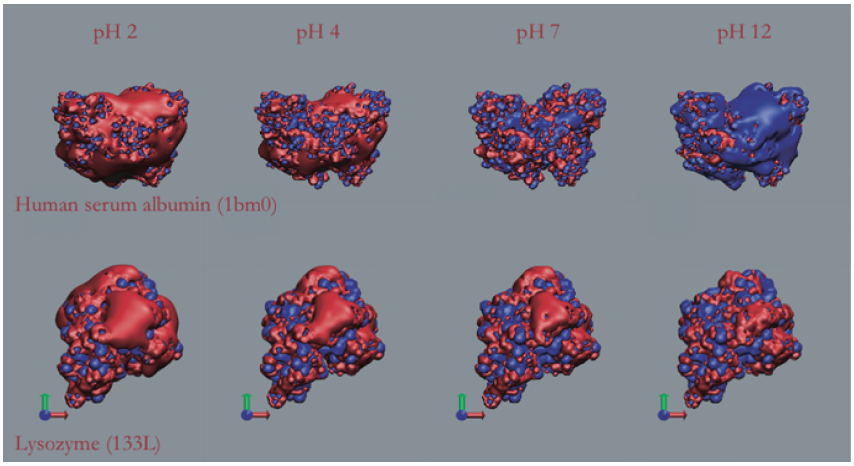


Figure 1.19 Distribution of positively charged (red) and negatively charged (blue) residues on the surface of lysozyme and human serum albumin.

A histogram showing the distribution of the pIs of many common proteins is given in Figure 1.18. As can be seen in this graph, the majority of the proteins have a slightly acidic isoelectric point and this is indeed found for the proteins present in many microorganisms such as *E. coli*. As a result, it is often easier to purify alkaline proteins that can be adsorbed on cation exchangers, because most of the host cell proteins are unlikely to be retained and will pass through these resins. Many monoclonal antibodies have high isoelectric points, allowing the development of platform purification processes using cation exchangers.

A final important consideration with regard to protein charge is the spatial distribution of the charged residues. An example illustrating the location of positive and negative charges on the surface of lysozyme or human serum albumin at neutral pH is shown in Figure 1.19. A consequence of this heterogeneous spatial distribution is that frequently the net charge of the protein is not sufficient to determine whether the protein will bind or not to an oppositely charged surface. For example, as a result of the localized concentration of positively-charged residues, it is possible for a protein to bind to a cation exchanger at pHs well above the protein pI, where the net charge is highly negative, or, vice versa, for a protein to bind to an anion exchanger at pHs well below the protein pI, because of localized negatively-charged residues.

1.2.2.4 **Hydrophobicity**

The hydrophobicity of a protein is determined by the side chains of its non-polar amino acids. Although the term hydrophobicity is commonly used, a precise definition is difficult and is extensively debated. The transfer of an apolar compound into a polar liquid such as water is associated with heat and quantified as free energy. The hydrophobic effect is strongest when entropic effects are dominant. Hydrophobic effects increase with the surface tension of water which is due to the

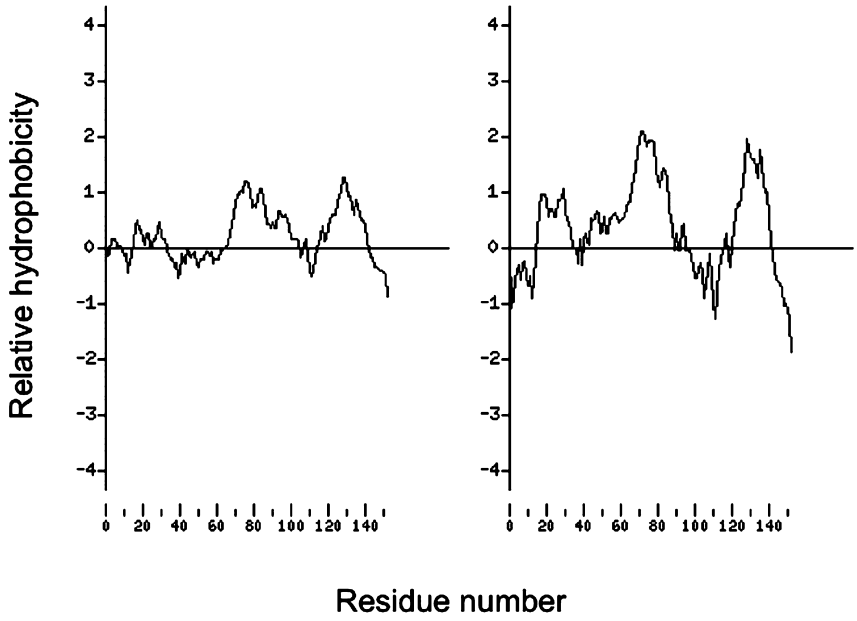


Figure 1.20 Calculated hydrophobicity of human Cu-Zn superoxide dismutase calculated according to different hydrophobicity scales: Hopp and Wood [22] (left) and Kyte and Doolittle [23] (right).

attraction between the molecules in the liquid caused by various intermolecular forces. Hydrophobic effects are thus mainly due to the strong hydrogen bonds in water, while van der Waals forces generally play a minor role.

The hydrophobicity of proteins can be theoretically calculated from the transfer energy of amino acids from an apolar solvent into water (see Figure 1.20). In a peptide chain the α-amino group and the carboxyl group are absent because they have reacted to form peptide bonds. Thus the free energy of transfer of amino acids does not totally reflect the hydrophobicity of a protein and strongly depends on the hydrophobicity scales which are used for calculating the hydrophobicity of the protein.

The distribution of surface-exposed hydrophobic residues in proteins is not homogenous. This is illustrated for lysozyme in Figure 1.21. The density and distribution of these residues at the surface of a proteins is the basis for *hydrophobic interaction chromatography* (HIC) where surface hydrophobic residues interact with a mildly hydrophobic matrix. Since incorrect folding may lead to variations in the number of surface exposed-hydrophobic residues, HIC may be used as a tool to separate native proteins from misfolded isoforms.

Another approach to measuring the hydrophobicity of a protein is by measuring its retention in a chromatography column packed with a hydrophobic medium. In this case, if the protein does not unfold, the retention is related to the number of hydrophobic amino acid side chains exposed at the surface [24]. The hydrophobic-

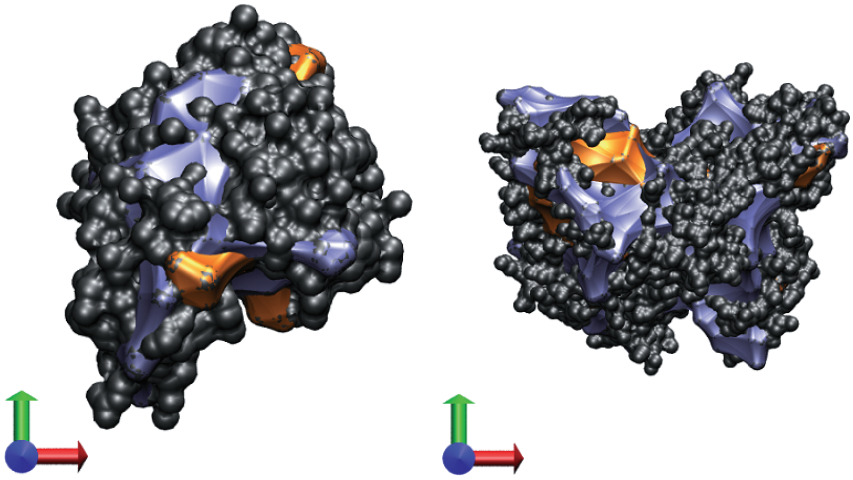


Figure 1.21 Distribution of hydrophobic clusters at the surface of lysozyme (left) and human serum albumin (right) at neutral pH. Yellow and light blue indicate hydrophobic and hydrophilic patches, respectively.

ity obtained by this method is relative and depends on the applied methodology, so it is useful only for ranking purposes.

1.2.2.5 **Solubility**

Solubility is often a critical consideration in downstream processing, since it can vary dramatically with pH, ionic strength, and salt type. Predicting the solubility of a protein in aqueous media from its structure is difficult and empirical measurements are usually needed. Protein solubility varies dramatically. Some proteins, e.g. Cu-Zn superoxide dismutase, have solubility as high as 400 mg ml^{-1} while others, e.g. recombinant interferon-γ, are soluble at concentrations less than 10 mg ml^{-1}. In general, protein solubility is lowest at the isoelectric point, where the net charge is zero, but varies with ionic strength, which is defined as follows

$$I = \frac{1}{2}\sum_{j=1}^{n} c_j z_j^2 \tag{1.16}$$

where c_j is the concentration of ion j and z_j its charge. The solubility of β-lactoglobulin as a function of salt concentration and pH, is shown as an example in Figure 1.22.

In general, salts at low concentrations increase the solubility of a protein, a process referred to as 'salting in'. Conversely, at high concentrations salts reduce protein solubility, which is referred to as 'salting out'. The magnitude of these effects is however highly dependent on the type of salt, as shown for example in Figure 1.23 for carboxyhemoglobin.

Protein solubility trends can be described by the extended form of the Debye–Hückel theory. Accordingly, we have:

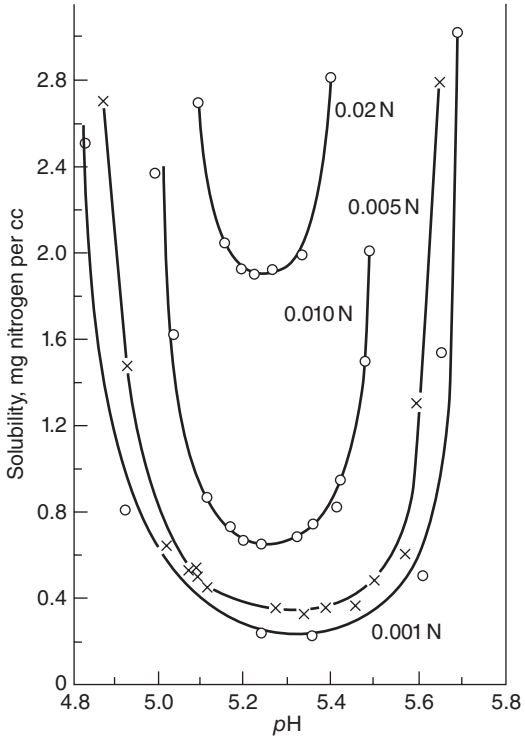


Figure 1.22 Solubility of β-lactoglobulin as a function of pH at four different concentrations of sodium chloride. Reproduced from [25].

$$\log\frac{w}{w_0} = \frac{0.5 \cdot z_1 \cdot z_2\sqrt{I}}{1 + A\sqrt{I}} - \kappa_s I \tag{1.17}$$

where w is the protein solubility in the actual solution, w_0 the solubility of the protein in water, z_1 and z_2 the salt charges, and κ_s and A are salt- and protein-specific empirical parameters. At high ionic strengths, Equation 1.17 reduces to the following log-linear relationship:

$$\log\frac{w}{w_0} = \beta - \kappa_s I \tag{1.18}$$

which is shown for various proteins in Figure 1.24.

The effect of the type of salt on protein solubility was formally described for the first time by the Hofmeister [28] who ranked the anions and cations according to their ability to precipitate proteins, which is generally known as the Hofmeister or lyotropic series:

$$\text{Anions: } SO_4^{2-} > Cl^- > Br^- > NO^{3-} > ClO^{-4} > I^- > SCN^-$$

$$\text{Cations: } Mg^{2+} > Li^+ > Na^+ > K^+ > NH_4^+$$

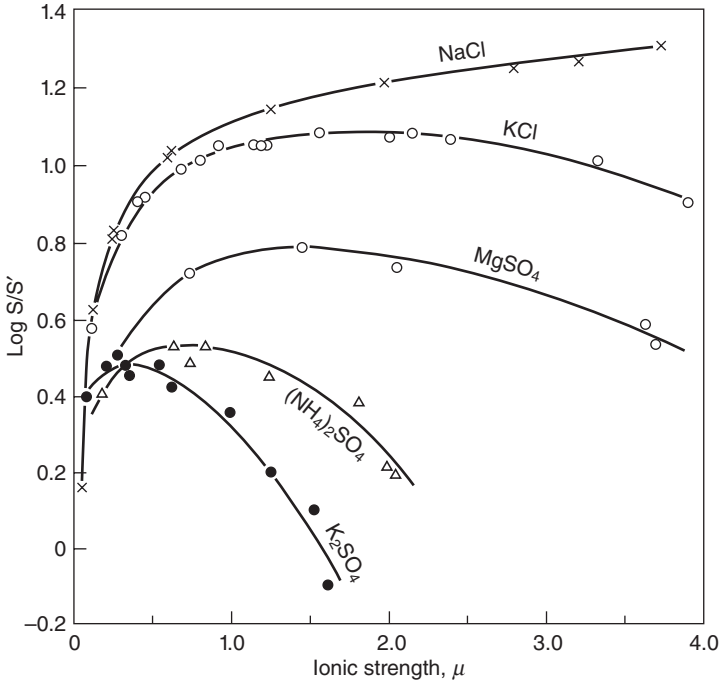


Figure 1.23 Solubility of carboxyhemoglobin in aqueous solution with different electrolytes at 25 °C. S and S^0 are used in lieu of w and w_0. Reproduced from [26].

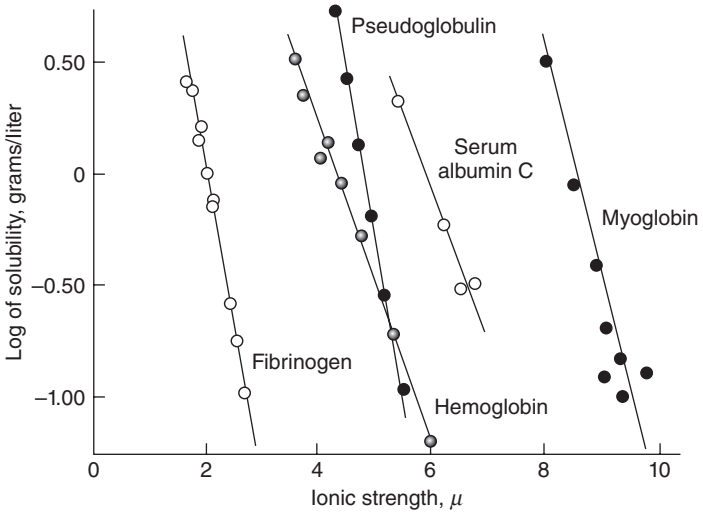


Figure 1.24 The solubility of proteins in ammonium sulfate solutions. Reproduced from [27].

A simple interpretation of this series is that certain ions bind free water decreasing the ability of the protein to remain in solution. Interestingly the salts in the Hofmeister series also correlate with the so-called Jones–Doyle B-coefficient and the entropy of hydration so that both appear to be related to the effects of salts on the structure of water.

Finally, it should be noted that in practice, the selection of salts for use in downstream processing depends not only on the Hofmeister series, but also on factors such as price, availability, biocompatibility, and disposal costs.

1.2.2.6 **Stability**

Two different types of stability need to be considered for proteins: the conformational or thermodynamic stability, and the kinetic or colloidal stability. The conformational stability of a protein is described by the free energy ΔG of the equilibrium between native and the unfolded states. The transition of the native folded form, *N*, into the unfolded form, *U*, is described by the following quasi-chemical reaction:

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} U \tag{1.19}$$

where k_1 and k_{-1} are rate constants. The corresponding equilibrium constant $K_{eq} = [U]/[N]$ is usually very low in aqueous solution, since protein folding is generally thermodynamically favored as a result of the concentration of the hydrophobic residues in the protein core. The corresponding ΔG is given by the following equation

$$\Delta G = -RT \ln K_{eq} \tag{1.20}$$

Representative values are given in Table 1.8 along with the corresponding 'melting temperature', which is defined as the temperature at which half of the protein is in the unfolded state. Kosmotropic (or cosmotropic) salts and polyols such as sorbitol or sucrose stabilize proteins while chaotropic salts or urea at higher concentrations have a destabilizing effect on protein conformation.

Kinetic stability, on the other hand, can be described by the following equation:

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} U \xrightarrow{k_2} A \tag{1.21}$$

which shows a further kinetically-driven step from the unfolded state to an irreversibly aggregated state *A*. Proteins with a high k_2 exhibit a low kinetic stability. The overall stability thus depends on both thermodynamic and kinetic effects. It is possible, for example, for an added salt to decrease kinetic stability, while enhancing overall stability as a result of thermodynamic effects. However, this effect is often difficult to predict, so that it in practice, overall stability and shelf-life are measured empirically [30].

Table 1.8 Thermodynamic stabilities of proteins. Data for chymotrypsinogen from [29].

Protein	Conditions	Free energy of the unfolding reaction ΔG, kcal mol^{-1}	Melting temperature °C
Horse Cytochrome c at pH 6 and 25 °C	0 M urea	31.3	n.a.
	2 M Urea	22.3	n.a.
	4 M Urea	14.2	n.a.
	6 M Urea	3.2	n.a.
Hen egg white Lysozyme at pH 3.0	24 °C	41.0	n.a.
	40 °C	30.4	n.a.
	55 °C	14.7	n.a.
	75 °C	–5.9	n.a.
Bovine chymotrypsinogen at melting temperature and pH 2.0	0 M sorbitol	0.015	42.9
	0.5 M sorbitol	0.146	44.9
	1.0 M sorbitol	0.235	44.2

1.2.2.7 **Viscosity**

Many of the solutions and suspensions encountered in bioprocessing are highly viscous. This is especially true for fermentation broths that contain DNA and for highly concentrated protein solutions. In general, the viscosity, η, is related to the shear stress, τ, and the shear rate, $\dot{\gamma}$, by the following equation:

$$\tau = \eta \times \dot{\gamma} \tag{1.22}$$

For Newtonian fluids, η is a constant and the relationship between shear stress and shear rate is linear. For non-Newtonian fluids, however, η varies with shear rate and the relationship is non-linear. For example, the behavior of pseudoplastic fluids is described by the following equation:

$$\tau = K \times (\dot{\gamma})^n \tag{1.23}$$

where K and n are the consistency and flow index, respectively. For highly concentrated protein solutions and for many culture supernatants, n is smaller than unity, indicating that the apparent viscosity, $\eta = \tau/\dot{\gamma}$, decreases with increasing shear rate. The ranges of shear rates for the various solutions and suspensions encountered in bioprocessing are shown in Table 1.9.

Typical viscosities encountered in bioprocessing are shown in Table 1.10. In general, cell culture supernatants have viscosities lower than 10 mPa s, while cell homogenates are much more viscous with η-values of up to 40 mPa s. The greatest contribution to the viscosity of raw fermentation broths is DNA. Fortunately, however, both genomic and plasmid DNA are very sensitive to shear and are often mechanically degraded early on in the downstream process. DNAse enzymes, either naturally occurring or added intentionally, also help to degrade these molecules, thereby reducing viscosity.

Table 1.9 Typical shear rates encountered in bioprocessing.

Operation	$\dot{\gamma}$ (s^{-1})
Expanded bed	<10
Packed bed	$<10^3$
Stirred tank	10^2–10^3
High pressure homogenizer	10^6

Table 1.10 Viscosities of various fluids at 20 °C.

Liquid	Apparent viscosity mPa s
Water	1
Glycerol	1070
Ethanol	1.20
Acetonitrile	0.34
Clarified cell culture supernatant	<5
Blood	10
E. coli homogenate	<40
E. coli broth	<20
Penicillium chrysogenum fermentation broth	40 000
Heinz ketchup	50 000–70 000

The intrinsic viscosity $[\eta]$ is a measure of the contribution of a solute to the viscosity of a solution and is defined as:

$$[\eta] = \lim_{c \to 0} \frac{\eta - \eta_0}{\eta_0 c} \tag{1.24}$$

where η_0 is the viscosity in the absence of the solute and c the solute concentration. At a semi-dilute limit, η can be described as a function of c by the following polynomial expression:

$$\frac{\eta - \eta_0}{\eta_0} = [\eta]c + k_1[\eta]^2 c^2 + k_2[\eta]^3 c^3 + \ldots \tag{1.25}$$

At very high protein concentrations, however, semi-empirical models are needed [31]. As an example, Figure 1.25 shows the relative viscosity (η/η_0) of IgG solutions as a function of IgG concentration. For concentrations lower than about 100 mg ml^{-1} the data conform approximately to Equation 1.25. However, at higher concentrations the viscosity increases exponentially [31].

Table 1.11 lists the intrinsic viscosities of representative biomolecules. As can be surmised from these data, the intrinsic viscosity depends on the shape of the molecule. For instance, rod-shaped proteins have a higher intrinsic viscosity than

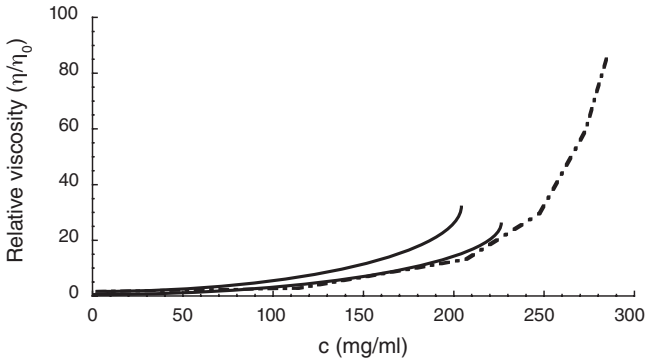


Figure 1.25 Viscosity of human, bovine, and pig IgG solutions as a function of IgG concentration. Data from [31].

Table 1.11 Intrinsic viscosities of various biologically important macromolecules in dilute solutions [32].

Shape	Substance	Molecular mass	$[\eta]$ (ml g^{-1})
Globular	Ribonuclease	13 680	3.4
	Serum albumin	67 500	3.7
	Ribosomes (*E. coli*)	900 000	8.1
	Bushy stunt virus	10 700 000	3.4
Random coils (unfolded proteins)	Insulin	2970	6.1
	Ribonuclease	13 680	16
	Serum albumin	68 000	52
	Myosin subunit	197 000	93
Rods	Fibrinogen	330 000	27
	Myosin	440 000	217
	Calf thymus DNA	15 000 000	>10 000

globular forms. An empirical relationship between $[\eta]$ and molecular mass M_r is given by the Mark–Houwink equation:

$$[\eta] = K(M_r)^a \tag{1.26}$$

where a is a parameter related to the 'stiffness' of the polymer chains. Theoretically, $a = 2$ for rigid rods, 1 for coils, and 0 for hard spheres. Empirically, however, values of $a = 0.6$, 0.7, and 0.5 have been found for BSA, ovalbumin, and lysozyme, respectively. Literature data (e.g. [33]) suggests a general relationship between intrinsic viscosity and the number of amino acid residues, which can be expressed as follows

$$[\eta] = 0.732\, n^{0.656} \tag{1.27}$$

with $[\eta]$ in $\mathrm{ml\,g^{-1}}$. Accordingly, larger proteins and protein aggregates have higher intrinsic viscosity than smaller proteins and monomeric forms.

1.2.2.8 Diffusivity

The molecular diffusion coefficient or diffusivity in solution, D_0, is a function of the size of the solute, the viscosity of the solution, and temperature. As previously noted, the Stokes–Einstein equation describes this relationship as:

$$\frac{D_0\eta}{T} = \frac{k_b}{6\pi r_h} \tag{1.28}$$

where k_b is the Boltzmann constant and r_h is the solute hydrodynamic radius. The diffusivities encountered in bioprocessing range widely from $1 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ for salts and other small molecules to $1 \times 10^{-9}\,\mathrm{cm^2\,s^{-1}}$ for large biomolecules such as DNA. Protein diffusivities in dilute aqueous solution are generally in the range of 10^{-6} to $10^{-7}\,\mathrm{cm^2\,s^{-1}}$. Table 1.12 provides a summary of typical diffusivities in dilute solutions at room temperature.

In general, protein diffusivities are 10–100 times lower than those of small molecules. Plasmids have even smaller values by a factor of as much as 1000. Figure 1.26 illustrates the effect of molecular mass on the diffusivity in dilute aqueous solution at room temperature. Tyn and Gusek [34] have provided the following correlation for globular proteins:

$$\frac{D_0\eta}{T} = \frac{9.2 \cdot 10^{-8}}{(M_r)^{1/3}} \tag{1.29}$$

where D_0 is in $\mathrm{cm^2\,s^{-1}}$, η in mPa s, T in K, and M_r in Da, which has an accuracy of ±10%.

Several approaches are available for the experimental determination of diffusivity and are reviewed, for example, by Cussler [35]. Commonly used approaches for

Table 1.12 Diffusivities in dilute solution at room temperature.

Solute	Solvent	Viscosity, η (mPa s)	Diffusivity, D_0 ($10^{-5}\,\mathrm{cm^2\,s^{-1}}$)
Benzoic acid	Water	1	1.00
Valine	Water	1	0.83
Sucrose	Water	1	0.53
Water	Ethanol	1.1	1.24
Water	Glycol	20	0.18
Water	Glycerol	>120	0.013
Ribonuclease (M_r = 14 kDa)	Water	1	0.12
Albumin (M_r = 65 kDa)	Water	1	0.060
IgG (M_r = 165 kDa)	Water	1	0.037
pDNA (M_r = 3234 kDa)	Water	1	0.004

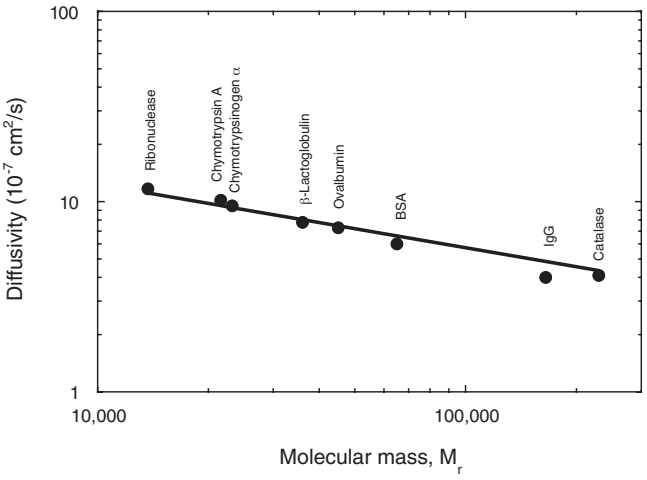


Figure 1.26 Diffusivity of globular proteins in dilute aqueous solution at room temperature. Data from [34].

proteins include dynamic light scattering (see Section 1.1.2.2), diffusion cells, Taylor dispersion-based methods, and microinterferometry.

1.3 Bioprocesses

This chapter discusses commonly used expression systems and the general structure of downstream processes needed to achieve the desired product purity. Special emphasis is placed on the production of recombinant proteins by fermentation and cell culture, which play a major role in industrial biotechnology.

1.3.1 Expression Systems

Many different expression systems have been developed for recombinant proteins, ranging from very simple bacteria to plants and animals. However, the number of host cells actually used in the industrial production of biopharmaceutical proteins is quite limited. The most popular bacterial strain is *E. coli*, BL21, which is used for the production of proteins whose biological activity does not require post-translational modifications. Protein expression in *E. coli* can occur in three different ways. The protein can be secreted into the periplasm, which is the space between cell membrane and cell wall; it can be expressed in the cytoplasm as soluble protein; or it can accumulate in the cell as inclusion bodies. Each system is effective for certain proteins. The cytoplasm of *E. coli* is a strongly reducing environment which hinders the formation disulfide bridges, whereas the

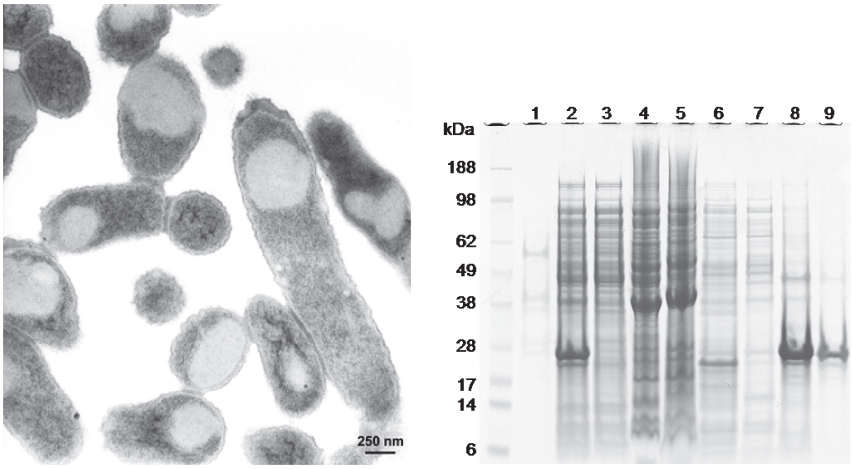


Figure 1.27 Left: Scanning electron micrograph of *E. coli* cells over-expressing a recombinant protein as inclusion bodies. Right: SDS-PAGE under non-reducing conditions. Lane 1: culture supernatant; lane 2: homogenate; lane 3: soluble fraction; lanes 4–7: supernatants from the wash steps; lanes 8–9: insoluble fraction containing the product.

periplasm offers a more oxidizing environment which allows the formation of these bridges thus facilitating folding. Some proteins which are toxic to the cells or possess a short half-life in the cytoplasm or periplasm have been successfully produced as inclusion bodies. In this case, the protein forms aggregates that cannot be attacked by proteases. On the other hand, although often not fully denatured, proteins expressed as inclusion bodies are generally not in their native conformation. Thus, a refolding procedure is typically required to generate a fully active form. Although refolding can be costly, on balance this approach is frequently economically viable, since expression levels in the inclusion body system are extremely high and simple washing procedures can be used to remove most of the host cell proteins resulting in relatively high initial purities. Figure 1.27 shows an example of *E. coli* cells over-expressing a protein as inclusion bodies and the corresponding SDS-PAGE analysis at various stages in the process.

The yeast cells *Saccharomyces cerevisae* and *Pichia pastoris* have been successfully used for over-expression of various recombinant proteins including insulin and albumin. *Saccharomyces cerevisiae* is also used for the production of the hepatitis surface antigen. However, mammalian cells are used to produce the majority of biopharmaceutical proteins. Although mammalian cell culture is generally more complex, these cells can carry out complex post-translational modifications, such as glycosylation, which are often critical to proper biological and pharmacological activity. *Chinese Hamster Ovary* (CHO) cells are the most commonly used mammalian expression system, especially for recombinant antibodies. The human cell line PerC6 has been developed more recently and is also used for production of

some recombinant proteins. CHO and PerC6 are able to over-express antibodies in concentrations as high as 15 mg ml^{-1}. Such high product titers are achieved mainly because the expression of proteins in these systems is generally independent of growth. As a result, the cells can be maintained in a bioreactor for long periods in a viable productive state and can be cultivated to very high densities, up to 10^8 cells per ml.

Other mammalian cell lines used in recombinant protein production include *baby hamster kidney* cells (BHK), Vero cells, and *Madin–Darbyn canine kidney* cells (MDCK). Several coagulation factors that require γ-carboxylation are produced in BHK. Vero cells, derived from monkey kidneys, and MDCK cells are used for the production of vaccines. Insect cells such as cells from pupal ovarian tissue of the fall armyworm *Spodoptera frugiperda* (Sf9), have also been proposed for over-expression of recombinant proteins. These cells can be infected with insect viruses such as the Autograph California nuclear polyhedrosis virus also known as baculo virus. Although easy to handle, so far insect cell systems have not been licensed for the industrial production of biopharmaceutical proteins.

In the past, transgenic animals were considered to be excellent production systems, since proteins can be secreted into the milk with titers up to 5 mg ml^{-1}. A decade ago such titers could not be achieved with mammalian cell culture. Thus, at that time these expression systems were often preferred. Modern advances in cell culture, however, have made it possible to routinely achieve even higher titers. Transgenic systems require the very tedious procedure of developing offspring with the appropriate expression characteristics. Purification of proteins from milk can also be cumbersome, especially when removal of casein by acid-induced precipitation is not compatible with the protein product. Furthermore, as an 'open production system', transgenic animals may be prone to safety issues and external contamination. The advantages of cultivation of mammalian cells in closed bioreactor systems together with titers higher than those achieved in transgenic milk and with simpler downstream processing are the reasons why cell culture is generally favored in the biopharmaceutical industry.

Plants and plant cells are also potential candidates for expression of proteins. Although different from that in mammalian cells, post-translational modifications are also possible in these organisms. However, expression in leaves, stems, or seeds presents significant recovery and purification challenges since the tissue or seeds must be ground or pressed and extracted which yields very complex mixtures. An interesting expression system is rhizo-secretion, where the protein is secreted into a cultivation fluid from hairy roots. Plant cells, on the other hand, present fewer downstream processing difficulties as these cells grow in very simple media. Another interesting system is the so-called 'olesin technology'. In this case, lipids are accumulated in plant cells in the form of small droplets. Proteins, known as olesins can be concentrated as a monolayer on the surface of these droplets. After fusing a recombinant protein to the olesins, primary recovery of the target protein can be achieved by the relatively simple process of harvesting the oil droplets.

1.3.2 Host Cells Composition

The composition of the host cells has important effects on downstream processing when the product is expressed intracellularly, since in this case the cellular components are the major impurities. However, host cell components are also found extensively as impurities in secreted products, since cell lysis always occurs to some extent during fermentation. In fact, in some cases, cultivation procedures that yield high titers, such as those used for antibody production, result in partial lysis of the cells, which in turn causes contamination of the product with host cell components. An overview of the composition and physical characteristics of the major host cells is given in Tables 1.13 and 1.14.

As can be seen from these tables, mammalian cells contain more protein but less nucleic acid than bacteria. Bacteria and yeast also contain numerous cell wall components. These components are frequently insoluble and are efficiently removed during early processing steps. These early steps are significantly affected by the cell density and viscosity of the broth. Extremely high cell densities can be obtained for yeast, particularly for *P. pastoris*, for which cell densities up to 400 g of cells per liter have been reported. Such suspensions are extremely difficult to clarify; the suspension must often be diluted in order to centrifuge the broth. Nucleic acids are present in the form of DNA and various forms of RNA. These compounds are responsible for the high viscosity of the broth but are often rapidly degraded by mechanical shear or endogenous nucleases.

A final consideration is the mechanical stability of the host cells. Bacteria and yeast cells are generally mechanically very stable and shear rates above $10^6\,s^{-1}$ are necessary to break these cells. Such high shear rates can only be attained by using special equipment such high-pressure homogenizers or French presses. By comparison, mammalian cells are much weaker. The burst force needed to destroy a yeast cell is in the range of 90 μN whereas the burst force for mammalian cells is in the range of 2–4 μN. The burst force increases with culture length in a batch

Table 1.13 Composition of common host cells for expression of recombinant proteins.

Organism	Composition (% dry weight)			Cell count per ml	Dry mass/wet mass ($mg\,ml^{-1}$)
	Protein	Nucleic acids	Lipids		
E. coli	50	45	1	10^{11}	20/100
Yeast	50	10	6	10^{10}	80/400[a)] 130/400[b)]
Filamentous fungi	50	3	10	10^9	100/400
Mammalian cells	75	12	Up to 10	10^7–10^8	0.17 to 1.7/1 to 10

a) For a high density culture of *P. pastoris* grown on glucose medium and b) grown on methanol.

Table 1.14 Composition of single cells for expression of recombinant proteins.

Component	Amount per HeLa cell	Amount per *E. coli* cell
Total DNA	15 pg[a)]	0.17 pg[b)]
Total RNA	30 pg	0.10 pg
Total protein	300 pg (5×10^7 molecules with an approx. M_r of 40 000)	0.2 pg (3×10^6 molecules with an approx. M_r of 40 000)
Dry weight	400 pg	0.4 pg
Wet weight	2500 pg	2 pg
Diameter	18 µm	0.5 × 3 µm
Volume	4×10^{-9} cm^3	

a) HeLa cells are hypotetraploid and contain four copies of each chromosome. The DNA content of normal cells is approx. 5 pg/cell.
b) A fast growing *E. coli* cell contains on average a four-fold repetition of its genome. The weight of each genomic DNA molecule is about 0.0044 pg.

culture, which is consistent with the observation that older cells are more difficult to disrupt.

1.3.3 Culture Media

Modern biopharmaceuticals are commonly produced with so-called defined media whose components are chemically characterized. In the past, yeast, meat, and soy extracts, produced by proteolytic degradation and extraction, were commonly used for the cultivation of bacteria and yeast cells. The standardization of such raw material was extremely difficult resulting in substantial batch to batch variations. Similarly, until recently it was common to supplement cultivation media for mammalian cells with fetal calf serum, in concentrations up to 10%. In addition to added complexity and cost, such supplements can introduce undesirable adventitious agents, such as prions, which can significantly increase the challenges of downstream processing. Although testing for such agents may still be required, the use of defined media greatly simplifies downstream processing.

Media for the industrial cultivation of bacteria are usually very simple and provide the essential sources of carbon, nitrogen and phosphate. Examples are given in Table 1.15. Sometimes a cocktail of trace elements is added, but frequently the trace elements present in the water are sufficient.

When the fermentation pH is controlled by the addition of NaOH, conductivities as high as 40 mS cm^{-1} can be reached by the end of the cultivation period. Such high conductivities can interfere with downstream processing operations such as ion exchange, therefore dilution or diafiltration steps are required. This difficulty

Table 1.15 Composition of defined culture media for the cultivation of *E. coli*.

Compound	Concentration (mg ml^{-1})
Glucose	1.0
Na_2HPO_4	16.4
KH_2PO_4	1.5
$(NH_3)_2PO_4$	2.0
$MgSO_4.7H_2O$	0.2
$CaCl_2$	0.01
$FeSO_4.7H_2O$	0.0005

may be circumvented by using ammonia to control the pH, which typically results in lower conductivity of the culture supernatant.

In addition to the salts and sugar, which are required for cell metabolism, production of recombinant proteins in bacteria typically requires the addition of an inducer, most commonly *isopropyl-β-D-thiogalactopyranoside* (IPTG), which is used to activate protein expression when a certain cell density is reached. The use of natural compounds as inducers is advantageous since such species are readily degraded in the culture and are not significant impurities. On the other hand, detergents or oils added to the culture as anti-foaming agents, although present in relatively small amounts, can affect downstream processes since they tend to foul membranes and chromatography matrices.

Culture media for yeast are similar to those used for *E. coli*. Methanol is frequently used as the inducer for systems based on the *alcohol oxidase promoter* (AOX) expression system. Mammalian cell culture, on the other hand, requires much more complex media including glucose as the carbon source, amino acids, vitamins, inorganic salts, fatty acids, nucleotides, pyruvate, and butyrate. This basal medium is supplemented with proteins for oxygen transport, hormones, and growth factors. Oxygen transport proteins such as transferrin contain bound iron. In order to create a totally protein-free medium these proteins are often replaced by iron chelators such as ferric citrate, ferric iminodiaectic acid, ferric ammonium citrate and tropolone (2-hydroxy-2,4,6-cycloheptatrien-1-one). However, these compounds can also interfere with downstream processing. In addition, under slightly acidic conditions, ferric citrate forms a gel, which is difficult to separate from proteins and other biomacromolecules.

Several other additives may be present in cell culture media: pH indicators such as phenol red, added to laboratory scale culture media often bind to ion exchange resins and are best avoided for large-scale cultivation. However, polymers, such as poly(propylene glycol) or poly(ethylene glycol), are often needed in concentrations of up to 0.02% to protect the cells from shear stress. The pH in mammalian cell culture is typically regulated by the addition of CO_2, although high density cultures may require addition of NaOH. Final conductivities of less the 17 mS cm^{-1} are typical, making direct capture by ion exchange easier than capture from yeast and *E. coli* homogenates.

1.3.4
Components of the Culture Broth

In general, prior to harvest, the culture broth contains the following components: intact cells, debris from lysed cells, intracellular host cell components, unused media components, compounds secreted by the cell, and enzymatically or chemically converted media components. Oxygen is depleted since during primary recovery the oxygen supply is shut down and the residual dissolved oxygen is rapidly consumed. The low oxygen content can induce necrosis and cells may rapidly die and lyse during this phase. Some cell types begin autolysis after just 30 min without oxygen. As a result, depending on the cell type rapid separation of the cells from the broth supernatant is necessary to maintain a low level of host cell impurities. The high concentrations of CO_2 that may be produced by the residual cells in the culture broth will fragment and shift the pH towards the acidic region. CO_2 is much more soluble in aqueous solutions than oxygen, so that substantial concentrations may be present. Dissolved CO_2 is rapidly liberated when the pH is adjusted for downstream processing and forms bubbles which may then enter chromatography columns and disrupt the packing.

Intracellular host cell components appearing as impurities in a culture supernatant can be estimated from the following equation:

$$(1 - \text{Fraction of live cells}) \times (\text{Cell count}) \times (\text{Amount of intracellular component per cell})$$

where the concentration of intracellular components can be estimated from Tables 1.13 and 1.14. Thus, from a downstream processing perspective high cell viability is desirable; this not always possible however. For example, in high-titer antibody production by cell culture, the cells often lyse in the final stage of cultivation. As a result, supernatants from mammalian cells cultures grown using defined media will contain host cell proteins in the range of 1–3 mg ml^{-1} (1000 to 3000 ppm). These levels must be reduced to less then 100 ppm in the final product.

1.3.5
Product Quality Requirements

The manufacture of biological products for pharmaceutical applications must follow general guidelines that have been established by the regulatory framework. Three keywords summarize the principal product quality requirements: *purity*, *potency*, and *consistency*. Industrially, these requirements must be met by processes that are economically viable and bring products to the market rapidly. Downstream processes must be designed to obtain sufficient purity while maintaining the potency or pharmacological activity in a consistent manner.

1.3.5.1 Types of Impurities

Purity requirements for biopharmaceuticals vary depending on the particular application. Thus, it is not possible to specify absolute values. However, an important distinction can be made among the various impurities which are frequently

categorized as critical or non-critical. A non-critical impurity is an inert compound without biological relevance. This can be, for instance, residual PEG from an extraction process or a harmless host component such as a lipid. On the other hand, endotoxins or growth factors secreted into the culture supernatant are examples of critical impurities, since they exert adverse biological activity. These impurities need to be traced throughout the process. Impurities stem either from the cultivation process or from materials added for processing. Examples of process-related impurities are solubility enhancers, redox-buffers, enzyme inhibitors, and compounds leached from filters or chromatography media. Monomers which leak from the polymeric materials that come into contact with the product are sometimes of particular concern. Extensive testing and documentation of their removal is generally mandatory.

The term *adventitious agent* is used to describe potentially infectious impurities that have not been added intentionally and are not essential to the process but are typically extremely hazardous. They may enter the process as a result of contaminated raw materials or cells. Examples of adventitious agents are viruses, virus-like particles, and prions such as transforming spongiform encephalitic agents. Certain toxic chemicals may also be considered to be adventitious agents. Finally, bioburden, originating from microbial contamination from the air or personnel or from inadequately cleaned equipment, can also have serious effects and must be carefully monitored and controlled.

Table 1.16 summarizes the measures required to demonstrate the removal of adventitious agents and other impurities. This demonstration is usually carried out in scaled down experiments, because it would obviously be counterproductive to intentionally contaminate the production plant with an adventitious agent. For these determinations, also known as spiking experiments, a bolus of an adventitious agent, e.g. a virus, is added to the raw feed stream entering a purification process step. The virus titer before purification $a' = \log_{10}(\text{Feed titer})$ and after purification $a'' = \log_{10}(\text{Harvest titer})$ are determined and the *log-virus reduction* (*LVR*) factor is calculated as follows:

$$LVR = \log_{10}(\text{Feed titer}) - \log_{10}(\text{Harvest titer}) = a' - a'' \tag{1.30}$$

Table 1.16 Measures required to demonstrate the removal of adventitious agents and impurities.

Measures	Adventitious agents	Other impurity
Spiking experiments to demonstrate clearance	Yes	No
Starting material preferably free of contaminating agents	Yes	No
Clearance measured at each step	No	Yes
Control of final product	Yes	Yes

Table 1.17 Typical virus clearance values of murine leukemia virus for purification of a recombinant antibody. Data from [36].

Step	*LRV*
Capture	4.23
Low pH	>4.4
Cation exchange chromatography	2.13
Anion exchange chromatography	4.98
Hydrophobic interaction chromatography	4.19

In order to account for the effect of volume changes (e.g. a mere 1 : 10 dilution results in a *LVR* of 1), the following individual reduction factor, R_i, is also calculated:

$$R_i = LVR - \log_{10} \frac{V'}{V''} \tag{1.31}$$

where V' and V'' are the feed and harvest volumes, respectively. Finally, the *LVR* of the individual process steps are added together to arrive at a cumulative *LVR* for the entire process. Table 1.17 illustrates a typical virus clearance scheme for an antibody purification process.

Although almost all mammalian cells are infected with viruses making the validation of viral clearance obligatory, efforts are frequently made to omit these procedures by utilizing platform processes that have demonstrated clearance efficiency.

1.3.5.2 **Regulatory Aspects and Validation**

Biopharmaceutical product quality and process validation are subject to regulations by the individual governments. In the United States, the legal framework is published in the Code of Federal Regulations 21 (21 CFR), Subchapter F Biologics. The *US Food and Drug Administration* (US FDA) is responsible for its implementation. The US code defined biological products as follow: 'Biological product means any virus, therapeutic serum, toxin, antitoxin, or analogous product applicable to the prevention, treatment or cure of diseases or injuries of man'. Plasmids, cells and tissues are not explicitly mentioned, but are considered to be part of this definition. In the *European Union* (EU), the legal framework is still under the sovereignty of the individual member states, although an EU-wide umbrella organization, the European Medical Agency (EMEA), has been founded, with the goal of harmonizing the EU regulatory structure. The guidelines for medicinal products derived by biotechnology as set out by the Commission of the European Community, defines biologics as products derived from:

1) human blood, other human body fluids or human tissue
2) animal blood
3) microorganisms or components of microorganisms

4) animals or microorganisms for active or passive immunization
5) monoclonal antibodies
6) recombinant DNA products

The existence of multiple regulatory frameworks adds complexity to the biopharmaceutical industry. Thus, the *International Conference on Harmonization* (ICH) has been established to develop a common international regulatory framework. ICH brings together the regulatory authorities of Europe, Japan and the United States, as well as experts from the pharmaceutical industry in these three regions to discuss scientific and technical aspects of product registration. The purpose is to make recommendations on ways of achieving greater harmonization in the interpretation and application of technical guidelines and the requirements for product registration. While progress is being made in the harmonization process, the regulatory framework of each country remains in effect.

Over the years, the *Center for Biologics Evaluation and Research* (CBER) of the US FDA has issued detailed recommendations for the manufacture of biologicals in a series of 'Points to Consider' papers. The first of these papers was released in 1983 with the title 'Interferon Test Procedures: Points to Consider in the Production and Testing of Interferon Intended for Investigational Use in Humans'. Subsequent papers include:

1) Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology – 4/10/1985
2) Guidelines on Validation of the Limulus Amebocyte Lysate Test as an End-Product Endotoxin Test For Human and Animal Parenteral Drugs, Biological Products and Medical Devices – 12/1987
3) Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals – 1995
4) Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use – 2/28/1997
5) Guidance for Industry: Monoclonal Antibodies Used as Reagents in Drug Manufacturing – 3/29/2001

A complete list can be found on the official website of CBER: http://www.fda.gov/Cber/guidelines.htm#95.

Validation is a critical aspect of biopharmaceutical process development. According to existing ICH definitions, critical parameters that may influence product quality need to be validated. After validation, a *standard operating protocol* (SOP) is established, which describes the process and the allowed variations. *Critical operational parameters* are defined as a limited sub-set of process parameters that significantly affect critical product quality attributes when varied outside a meaningful, narrow (or difficult to control) operational range. Consider, for example, the operation of a chromatographic purification step. As shown in Table 1.18 this operation will require the definition of a number of operating conditions as inputs, which in turn will result in certain performance characteristics, defined as outputs.

In order to validate the process, the input parameters must be varied over suitable ranges and the output measured. The critical parameters are then defined

Table 1.18 Examples of input and output parameters in a chromatographic separation process for proteins.

	Parameter							
Input parameter	pH	Temperature	Ionic strength	Load	Flow rate	Column height (residence time)		
Output parameter (Performance)	Purity	Concentration	Stability	Yield	DNA content	Host cell protein content	Endotoxin content	Back pressure

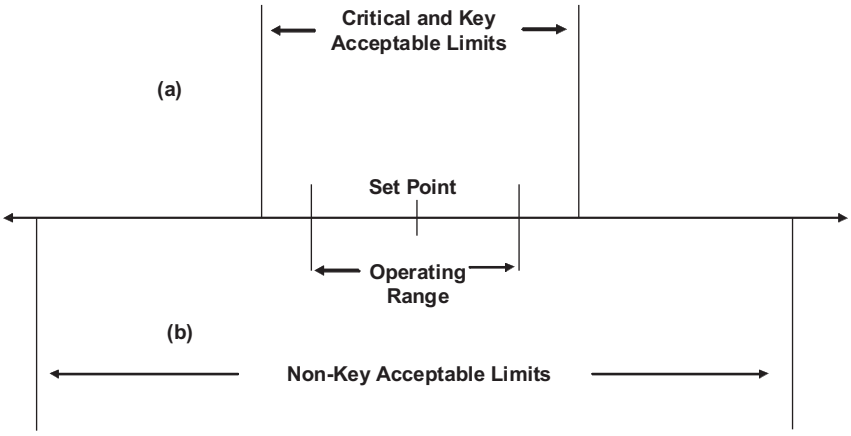


Figure 1.28 Definition of operation ranges for critical and non-critical parameters.

based on these experiments, which are usually carried out on a small scale. Suitably narrow operational ranges are established for these parameters as well as for non-critical parameters. Since the latter do not affect critical product quality attributes, their ranges will normally be broader than those for critical parameters (see Figure 1.28).

1.3.5.3 Purity Requirements

It is difficult to describe absolute purity requirements since they depend on the intended use of the biopharmaceutical, the dose, the risk–benefit ratio, etc. Table 1.19 provides only approximate values which are meant to serve as general guidelines.

Aggregates are an important concern for many biopharmaceutical proteins. It has been shown that aggregates can induce immune reactions or cause other side effects. Moreover, aggregates may constitute seeds for precipitation and reducing the shelf life of a product. As a result, controls on the percentage of dimers, oligomers and higher aggregate forms are often required. The leakage of ligands or

Table 1.19 General guidelines for purity, consistency, and potency of protein biopharmaceuticals.

Criteria	Requirement
Purity	
Specific protein content	>99.9%
Dimer/oligomer content	<1.0%
Ligand leakage	Usually < 1 ppm
Virus content	Absence with a probability of $<10^{-9}$
DNA content	<10 pg/dose
Endotoxin content	5 EU $kg^{-1} h^{-1}$
Prion content	Absence with a probability of $<10^{-9}$
Consistency	
Microheterogeneity	Permitted, but consistent
Impurities	Permitted, but consistent
Potency	
Folding	Correctly folded
Mutations	Correctly expressed, no mutations
Processing	Correctly processed

other leachable chemicals from chromatographic media and membranes is also an important concern since these materials can be immunogenic and toxic.

Viral contamination is obviously a critical issue as in the past it has been responsible for many iatrogenic diseases, such as those occurring because of contaminated blood products. Since absolutely complete removal of these adventitious agents is not possible, limits are often established on the basis of a risk–benefit analysis. For example, the *World Health Organization* (WHO) accepts for a vaccine one adverse case in 10^9 applications – hence, a probability value of 10^{-9} is suggested in Table 1.19. In practice, the difficulty of achieving such low probabilities is, of course, dependent on the dose. For example, the commonly set level of 10 pg of DNA per dose is relatively easy to achieve for a Hepatitis B vaccine, where one dose may contain only about 12 μg of protein. On the other hand, such a requirement may be rather difficult to meet in the case of recombinant antibodies, where one dose can consist of as much as 500 mg of protein.

As previously mentioned, most biopharmaceuticals are not fully defined individual molecular entities – rather, they consist of a large number of similar isoforms or variants (some recombinant antibody products contain as many as 200 identifiable variants). Since the biological and pharmacological activity can vary dramatically among different isoforms, it is important to maintain the distribution of these variants within established acceptable ranges. Because of the complexity of bioproduction systems, similar consistency must also be maintained for the impurity profiles as determined from analytical assays, in order to assure product safety. Finally, test systems must be established to control the potency *in vitro* and, where necessary, *in vivo*.

1.4
Role of Chromatography in Downstream Processing

Chromatography is the principal tool used for the purification of biopharmaceuticals. This can be explained by certain advantages of chromatography over other unit operations. First, chromatography provides very high separation efficiencies, which allow the resolution of complex mixtures with very similar molecular properties. Properly designed chromatography columns can have the separation efficiency of hundreds or even thousands of theoretical plates. By comparison, extraction and membrane filtration are usually limited to only a few stages. Second, chromatography columns packed with high capacity adsorbents are ideal for capturing molecules from the dilute solutions encountered in bioprocessing. In such systems, there is efficient contact between large volumes of solution and small amounts of the adsorbent packed into a column, resulting in either rapid concentration of the product or the nearly complete removal of contaminants present in small concentrations. By comparison, liquid–liquid extraction systems typically require similar volumes of the two phases in order to function properly, so that concentration is not very feasible. A further advantage is that chromatography can be undertaken in an almost closed system and the stationary phase can be easily regenerated. Finally, chromatographic methods are well established in many practical biopharmaceutical manufacturing processes and suitable equipment and packing materials are readily available. A perceived disadvantage of chromatography is the difficulty of scale-up within the constraints of the biopharmaceutical industry. However, as will be shown in the remaining chapters of this book, proper application of engineering tools in combination with adequate measurements allows the design of optimum columns for large-scale applications. Indeed, as shown recently by Kelley [2], chromatographic purification processes are suitable and technically and economically viable for protein purification on scales as high as 20 tons per year. Although there are no products which are currently made on such a large scale, the popularity of biopharmaceuticals is increasing rapidly so that such scales can be envisioned in the future.

Figure 1.29 illustrates the structure of a generic process for the recovery and purification of biologicals produced by microbial fermentation or animal cell culture. The initial steps where cells are separated are often referred to as *primary recovery*. This step requires different strategies dependent on whether the product is secreted into the culture medium or expressed in the cell, either as inclusion bodies, in soluble form in the cytosol or periplasm, or anchored to the membrane. Generally, chromatography plays a minor role in these initial steps, which are focused on the removal of suspended solids such as cells or cell debris. Sedimentation, centrifugation, deep bed filtration, and microfiltration or combinations thereof are normally used for these early steps. However, chromatography, implemented through the use of fluidized or expanded beds can also be used for the direct capture of secreted proteins from cell culture supernatants.

In these systems the liquid flows upwards through an initially settled bed of dense adsorbent particles. Above a certain flow velocity the bed expands and the

particles become fluidized allowing free passage of cells and other suspended matter while the product is directly captured by the adsorbent. This approach can be effective for dilute suspensions, but since bed expansion is directly influenced by the feed density and viscosity, the operation tends to be critically affected by variations in the composition of the broth. In practice, the high viscosity and cell density encountered in modern fermentation technology (up to 400 mg ml^{-1} wet cell mass for *P. pastoris* or 200 mg ml^{-1} for *E. coli*) make it difficult to implement this approach reliably on an industrial scale. An alternative possibility for early capture without clarification is to use adsorption beds packed with large particles, sometimes referred to as 'big beads'. If the particles are larger than about 400 μm in diameter, the interparticle spaces are sufficiently large to allow passage of cells and cell debris. Although the efficiency of capture is reduced by the diffusional limitations that accompany the larger particle diameter, the ensuing reduction in the number of processing steps can provide overall economic and operational advantages. Unlike expanded beds, packed bed processes are not very sensitive to feed viscosity so that reliable operation with large diameter beads can be achieved even with viscous feedstocks.

As seen in Figure 1.29, following primary recovery, the general downstream processing scheme consists of successive *capture*, *purification* and *polishing* steps, each comprising one or more operations. With only a handful of exceptions, current industrial processes for biopharmaceuticals almost exclusively employ chromatography for these three critical steps.

Figure 1.30 shows an example of the purification of a recombinant antibody. In this process, capture is realized using a selective adsorbent comprising of Staphylococcal Protein A immobilized in porous beads. This highly selective ligand

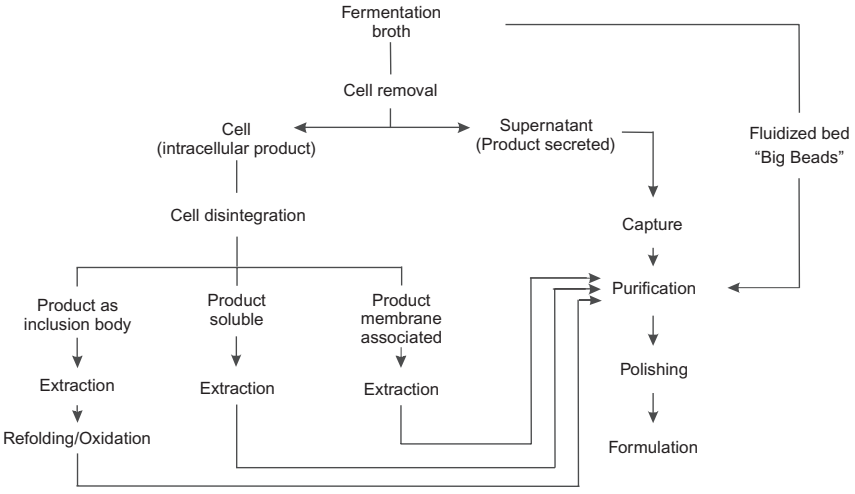


Figure 1.29 Generalized downstream processing flow sheet for purification of proteins starting with the unclarified fermentation broth.

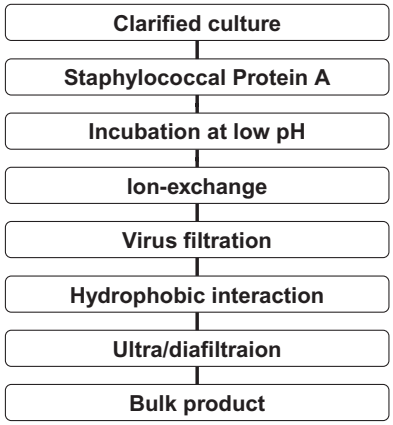


Figure 1.30 Generalized flow scheme for the purification of a recombinant antibody.

allows the direct loading of the clarified culture broth onto the capture column, and then selectively binds the antibody. In the subsequent steps, purification and polishing are conducted with ion exchange and hydrophobic interaction columns to remove host cell proteins and aberrant protein variants. Note that intermediate, non-chromatographic, steps are also included. Firstly, incubation at low pH and then a 'virus filtration' step are implemented for viral clearance. Secondly, an ultra/diafiltration step is included for buffer exchange and final formulation.

In addition to purification, chromatography is also employed in other bioprocesses. An important example is the use of chromatography to facilitate refolding of solubilized protein, which sometimes creates a bottleneck in industrial processes. Without simultaneous separation, misfolding and aggregation in particular, compete with the correct folding pathway. Aggregation may originate both from non-specific (hydrophobic) interactions of the predominantly unfolded polypeptide chains as well as from incorrect interactions of partially structured folding intermediates. As shown in Figure 1.31, aggregation reactions are second- (or higher) order processes, whereas correct folding is generally determined by first-order reactions [37].

In practice, refolding conditions (e.g. denaturant concentration) are adjusted so that the equilibrium distribution favors the formation of native protein (i.e. $k_2 \gg k_3$). The formation of intermediates is generally very fast, so that k_1 can be neglected. For the case where $k_3 \rightarrow 0$, we effectively have competing first and second (or higher) order reactions. Under these conditions, refolding in a batch system is described by the following equations:

$$\frac{d[U]}{dt} = -\left(k_2[U] + k_4[U]^n\right) \tag{1.32}$$

$$\frac{d[N]}{dt} = k_2[U] \tag{1.33}$$

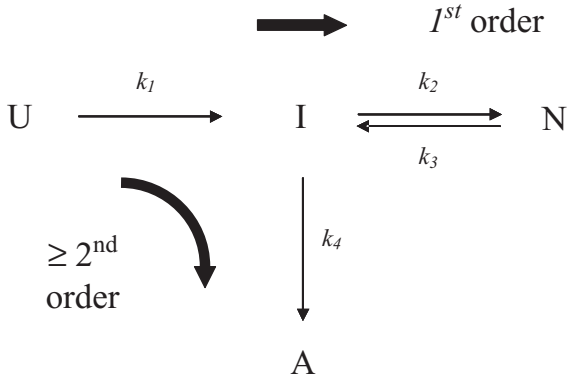


Figure 1.31 Simplified reaction scheme for protein refolding with aggregation of intermediates.

where the brackets denote concentrations, k_2 is the net rate constant of folding, k_4 the net rate constant of aggregation, and n is the reaction order. An analytical solution of these equations is available for $n = 2$ and is given by the following equation [37]:

$$Y(t) = \frac{k_2}{[U]_0 k_4} \ln\left\{1 + \frac{[U]_0 k_4}{k_2}\left(1 - e^{-k_2 t}\right)\right\} \tag{1.34}$$

where Y is the yield of the refolding reaction and $[U]_0$ is the initial concentration of unfolded protein. As time approaches infinity, the final yield of native protein is then given by the following equation:

$$Y(t \to \infty) = \frac{k_2}{[U]_0 k_4} \ln\left(1 + \frac{[U]_0 k_4}{k_2}\right) \tag{1.35}$$

This result suggests that dilution (i.e. low $[U]_0$) is a simple and effective way of ensuring high refolding yields. While this is effective and widely used in practice, the ensuing large solution volumes complicate further downstream processing and increase cost. Refolding in chromatographic columns also known as matrix-assisted refolding, can be a valuable refolding alternative to reduce the need for extensive dilution. The underlying mechanism leading to improved folding in chromatographic columns is not completely understood and may depend on the specific nature of the protein and the selected conditions. However, the effects can be dramatic as shown for example in Figure 1.32. In this case, refolding was conducted by separating the denaturing agent (urea) from the unfolded protein using size exclusion chromatography and allowing refolding to occur within the column. This resulted in a greater yield of folded protein compared to a simple dilution process. The apparent aggregation rate constant in this case was about 30 times smaller compared to that for the dilution process (Table 1.20). One possibility to explain this result is that aggregation may be inhibited within the matrix pores by

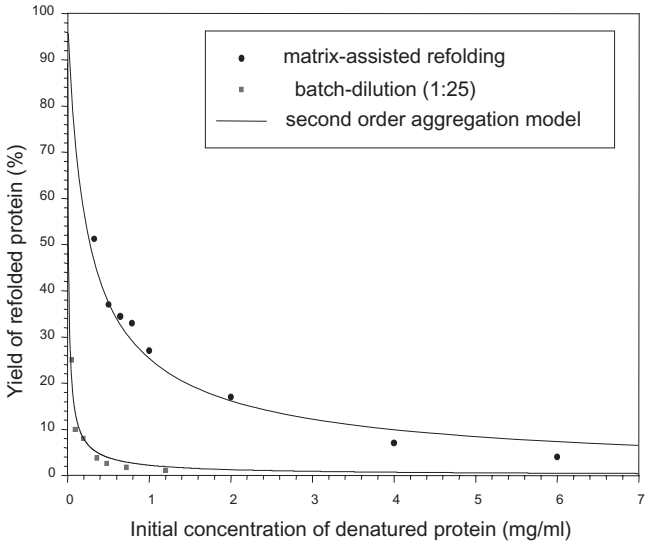


Figure 1.32 Refolding yield of a protein by batch dilution and with matrix-assisted refolding using size exclusion chromatography. Reproduced from [38].

Table 1.20 Refolding and aggregation rate constants of a protein in refolding by batch dilution and matrix-assisted refolding using size exclusion chromatography. Data from [38].

	Rate constants	
	Folding k_2 **(min^{-1})**	**Aggregation** k_4 **(ml mg^{-1}min^{-1})**
Batch-dilution	0.0012	0.3
Matrix-assisted refolding by SEC	0.0012	0.01

steric hindrance which allows a greater portion of the protein to follow the path toward correct folding.

In the example given in Figure 1.32 the unfolded protein was passed over a size exclusion column and the denaturant was slowly removed. Comparison of kinetic constants between conventional refolding by dilution into a refolding buffer and matrix-assisted refolding confirms that aggregation is suppressed in the column (Table 1.20).

SEC-promoted refolding is also possible in continuous processes, which can also include a recycling system for aggregated protein. Yield and productivity of a continuous refolding systems using *pressurized continuous annular chromatography* (P-CAC) considering initial protein concentration, residence time and recycling rate were extensively studied using α-lactalbumin as the model protein [39]. Also

countercurrent chromatography systems such as the *simulated moving bed* (SMB) can be used for continuous matrix-assisted refolding.

Ion exchange, affinity adsorption and hydrophobic interaction have also been used to facilitate refolding. A method based on the adsorption of the unfolded protein onto an ion exchange resin was introduced by Creighton [40]. Further improvements of this method include the application of more sophisticated buffer during loading and elution. The methods can be also executed in a continuous manner. The surface contact can initiate refolding. It has often been observed that final refolding takes place after the protein has been eluted from the column. Immobilization of denatured proteins on a solid support is often limited by the flexibility to regain the native configuration because of multi-point interactions with the matrix. Introduction of an N- or C-terminal poly-histidine-tag allowed the reversible one-point immobilization of the denatured protein on a solid support based on *immobilized metal affinity chromatography* (IMAC). Refolding can be achieved by a simple buffer exchange in a stepwise or gradient manner. The use of hydrophobic interaction chromatography (HIC) was described for the refolding of lysozyme, BSA, α-amylase and recombinant γ-interferon [41].

Immobilized folding catalysts and artificial chaperones have also been suggested as refolding aids. Mimicking *in vivo* folding systems was a further step in improving the yield of *in vitro* refolding. The chaperones or compounds mimicking chaperones are immobilized on a chromatography matrix. The protein solution is passed through such columns. The folded proteins are slightly retarded and the denaturant is exchanged. The immobilized chaperones prevent aggregation. Thus refolding can be achieved at higher concentrations or yield. It has to be borne in mind that a chaperone acts in a stoichiometric manner. Thus large amounts of chaperone protein are necessary to avoid aggregation.

References

1 Buchacher, A., and Iberer, G. (2006) *Biotechnol. J.*, **1**, 148.

2 Kelley, B. (2007) *Biotechnol. Progr.*, **23**, 995.

3 Aggarwal, S. (2008) *Nat. Biotechnol.*, **26**, 1227.

4 Rinaldi, A. (2008) *EMBO Rep.*, **9**, 1073.

5 Berg, J., Tymoczko, J., and Stryer, L. (2006) *Biochemistry*, Palgrave Macmillan.

6 Voet, D.J., and Voet, J.G. (2006) *Biochemistry, Textbook and Student Solutions Manual*, John Wiley & Sons, Inc., New York.

7 Egorov, T.A., Svenson, A., Ryden, L., and Carlsson, J. (1975) *Proc. Natl Acad. USA*, **72**, 3029.

8 Branden, C., and Tooze, J. (1991) *Introduction to Protein Structure*, Garland Publishing, Inc., New York.

9 Ueberbacher, R., Haimer, E., Hahn, R., and Jungbauer, A. (2008) *J. Chromatogr. A*, **1198**, 154.

10 Schmid, F.X. (1997) Optical spectroscopy to characterize protein conformation, in *Protein Structure A Practical Approach*, 2nd edn (ed. T.E. Creighton), IRL Press, Oxford, pp. 261–297.

11 Dayhoff, M.O. (1974) *Fed. Proc.*, **33**, 2314.

12 Nötling, B. (1999) *Protein Folding Kinetics*, Springer, Berlin.

13 Gokana, A., Winchenne, J.J., Ben-Ghanem, A., Ahaded, A., Cartron, J.P., and Lambin, P. (1997) *J. Chromatogr. A*, **791**, 109.

14 Kneuer, C. (2005) *DNA Pharmaceuticals* (ed. S. Martin), Wiley-VCH Verlag GmbH, Weinheim.

15 Petsch, D., and Anspach, F.B. (2000) *J. Biotechnol.*, **76**, 97.
16 Mach, H., Middaugh, C.R., and Lewis, R.V. (1992) *Anal. Biochem.*, **200**, 74.
17 Wilfinger, W.W., Mackey, K., and Chomczynski, P. (1997) *Biotechniques*, **22**, 474.
18 Tanford, C. (1976) *Biochemistry*, **15**, 3884.
19 Kuehner, D.E., Engmann, J., Fergg, F., Wernick, M., Blanch, H.W., and Prausnitz, J.M. (1999) *J. Phys. Chem. B*, **103**, 1368.
20 Tanford, C., and Roxby, R. (1972) *Biochemistry*, **11**, 2192.
21 Righetti, P.G., and Caravaggio, T. (1976) *J. Chromatogr. A*, **127**, 1.
22 Hopp, T.P., and Woods, K.R. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 3824.
23 Kyte, J., and Doolittle, R.F. (1982) *J. Mol. Biol.*, **157**, 105.
24 Mahn, A., Lienqueo, M.E., and Salgado, J.C. (2009) *J. Chromatogr. A*, **1216**, 1838.
25 Fox, S., and Foster, J.S. (1957) *Introduction to Protein Chemistry*, John Wiley & Sons, Inc., New York.
26 Green, A. (1932) *J. Biol. Chem.*, **95**, 47.
27 Cohn, E., and Edsall, J.T. (1943) *Proteins, Amino Acids, and Peptides*, Academic Press, New York.
28 Hofmeister, F. (1888) *Arch Exp. Pathol. Pharmakol.*, **24**, 247.
29 Gekko, K., and Timasheff, S.N. (1981) *Biochemistry*, **20**, 4677.
30 Ahrer, K., Buchacher, A., Iberer, G., and Jungbauer, A. (2006) *J. Biochem. Biophys. Methods*, **66**, 73.
31 Monkos, K., and Turczynski, B. (1999) *Int. J. Biol. Macromol*, **26**, 155.
32 Sibileva, M.A., Zatyaeva, A.A., and Matveeva, N.I. (2001) *Mol. Biol.*, **35**, 73.
33 Reisner, A.H., and Rowe, J. (1969) *Nature*, **222**, 558.
34 Tyn, M.T., and Gusek, T.W. (1990) *Biotechnol. Bioeng.*, **35**, 327.
35 Cussler, E.L. (1997) *Diffusion–Mass Transfer in Fluid Systems*, 2nd edn, Cambridge University Press, Cambridge.
36 Li, Y., and Galperina, O. (2005) PDA Viral Safety Workshop, Langen, Germany.
37 Kiefhaber, T., Rudolph, R., Kohler, H.H., and Buchner (1991) *J., Biotechnol.*, **9**, 825.
38 Schlegl, R., Necina, R., and Jungbauer, A. (2005) *Chem. Eng. Technol.*, **28**, 1375.
39 Schlegl, R., Iberer, G., Machold, C., Necina, R., and Jungbauer, A. (2003) *J. Chromatogr. A*, **1009**, 119.
40 Creighton, T.E. (1990) United States Patent No. 4977248.
41 Geng, X., and Chang, X. (1992) *J. Chromatogr. A*, **599**, 185.